THE ROLE OF STATISTICS IN THE DEVELOPMENT OF HEALTH CARE POLICY

Dorothy P. Rice, National Center for Health Statistics

The need for and the development of a National Health Policy has generated much discussion and debate in the United States in recent years. The consensus of public and private interests on the need for a health policy derives in large measure from the recent escalation of medical care expenditures. In fiscal year 1950, medical care expenditures amounted to $12 billion and represented 4.6 percent of the Gross National Product (GNP) and by Fiscal Year 1975, Health spending rose to $118 billion and consumed 8.3 percent of the GNP (Mueller). Although data for the year ending June 1976 are not yet available, the projections are for about $135 billion and 8.5 percent of GNP.

What is Health Care Policy?

Rising costs are a continuing and persistent characteristic of the health care industry. "The health sector bears the dubious distinction of having suffered from intractable inflation for years, long before the rest of the economy discovered it" (Russell). Coupled with the concern about inflation is the concern about the organization and provision of medical care services. In the 1960's, health care came to be perceived as a right rather than a privilege (Cooper, White). The enactment of the Social Security Amendments of 1965 that established the new public programs--Medicare and Medicaid--established a policy of providing coverage and care for two major population groups--the aged and the poor. That milestone legislation enunciated another policy, however. Its preamble prohibited Federal interference with "The practice of medicine or the manner in which medical services are provided."

The infusion of public dollars into the medical care system accompanied by the perception of the right to medical care but without regard to the shortcomings of our delivery systems, has placed a strain on the health care delivery system resulting in fragmentation of services, unequal access to health care and lack of continuity of care (Austin, Ball, Saward). These converging forces have produced a lively debate on the need to intervene in the organization and delivery of medical care services to effectively guide the use of the Nation's resources and fulfill the expectations and perceptions of the "right to health care."

A number of investigators have commented on the ability of medical care to favorably influence health. Their conclusions indicate that there is often a lack of positive correlation between the two (Abel-Smith, Fuchs, Illich). We obviously cannot do without medical care, for a cessation of all medical care would surely be followed by a worsening of health status. Nevertheless, questions are being raised and actively investigated as to whether increments or decrements in the amount of medical care result in improvement or deterioration in the level of health (Rice, Wennberg).

The recognition of a "health care crisis" undoubtedly had its roots in the escalation of medical care prices and expenditures in the late 1960's. The President's health message of 1971 and the White Paper "Towards a Comprehensive Health Policy for the 1970's" issued in May 1971 stressed the need for an "interlocking strategy" in solving the medical care problems facing the Nation. The annual "Forward Plan" for health issued in the past several years by the Assistant Secretary for Health also emphasized the multifaceted approach to health care policy. The National Health Planning and Resources Act of 1974 (Public Law 93-641) sets "equal access to quality health care at a reasonable cost" as a National priority. The first provision of this law calls for the issuance of National guidelines for health planning, which set forth National Health planning goals and standards relating to the appropriate supply, distribution and organization of health resources.

I have sketched this background in the development and emergency of health care policy in some detail to emphasize the recognition that health care policy is comprised of an inter-related and systematic approach to meeting the challenge of assurance of the right to health care. Thus, it must evolve from a set of policies that address the health care problems that face the Nation, including:

1. Escalating costs of health care and the concomitant burdens of financing medical services;

2. Variations in demand for and inequities in access to medical care services that partly arise from the maldistribution of medical care resources;

3. Uneven quality of medical care that emerges from the inappropriate application of medical knowledge, without regard to the balance between the hazards inherent in medical intervention and the benefits derived from it.

The Need for Data

The development of a health care policy or policies dictates the need for collection, analysis, and dissemination of data in all the above areas. The 1960's and 1970's have seen a considerable number of program developments that, on the one hand, have produced a tremendous amount of information; on the other hand, they created additional needs for both general purpose and program specific data essential to health management at every level of government.

Finagle's laws on information state: "The information you have is not what you want; the information you need is not what you can obtain" (Murnaghan). In spite of that pessimistic outlook on the fulfillment of data needs, I believe that the development of a coordinated, systematic, and responsive data base is essential and feasible for government to make informed decisions for policy development and planning, to assess the impact of these decisions, and to operate its programs effectively. Just as hard data, both economic (such as GNP and CPI) and social (health utilization and longevity data) have indicated the need and possible directions for developing a set of health care policies, so will statistical information be important in implementing and monitoring the policies and identifying areas for modifications.

Past experience has shown that the need for "hard data" is a function of our emerging modern society (Hauser). As society's demand for an improved standard of living brought about new social processes, the government increased its functions to deal with these expansions. This expansion will continue; new public policies of increasing sophistication can be expected (Kahn, Abelson). The development of these health policies will reflect this dynamic interaction between social change and governmental functions.

One of the most significant challenges is the development of data systems capable of satisfying the multiple needs for data and providing the data required for evaluation of the impact of health programs, while keeping respondent burden reasonable. The needs for data extend beyond the Federal government to State and Local governments and private providers and third party payors. They all need reliable, timely, and comparable data and analyses which describe the health status of the population (including statistical analyses of determinants of health and disease); the availability of resources; the accessibility and use of services, the costs of services and resources, the sources of funding, and the quality of care. The lack of such statistics severely limits our capacity to plan, manage and evaluate our tremendous investment in health resources and delivery systems. Clearly then, the development of a coordinated, systematic and responsive data base is essential.

Organization of Health Statistics Programs to Meet Data Needs

How is the health statistical community responding to the needs of policymakers? The history of Federal statistical activities goes back to the last century when information on births and deaths was collected periodically by the Bureau of the Census. This function was moved to the Public Health Service in 1946. The National Health Survey Act of 1956 authorized a continuing survey and special studies of sickness and disability in the U.S. population as well as methodological research. Since those days, the health statistical programs have been continually adapting to better meet the emerging needs of an expanding community of data users. These users are considerably more sophisticated than in the past. They demand considerably more information as well as more knowledge of the limitations of the information. We have to be especially careful not to measure the things that are easy to measure rather than those which would be helpful simply because those which would be more helpful cannot be measured well (Tukey)

My thesis is a simple one: We are developing a National and systematic approach to meeting the needs of policymakers at the Federal level. At the same time, we are cognizant of the State and local needs and have begun to address them. I would like to first briefly outline the developments in the Department of Health, Education, and Welfare.

Health Data Policy Committee

In recognition of the complexity of the health statistics systems and the widespread distribution of responsibility within DHEW, a Health Data Policy Committee was established in April 1974 to advise the Assistant Secretary for Health on current and long-term data needs for planning and management, on policies and procedures for coordination of health statistics activities, on proposals for major new health statistics systems, and on uniform minimum basic data sets.

The Committee has developed into an effective vehicle for a coordinated approach to data oriented health policy issues. It provides an internal forum for discussing and resolving specific issues and a focus within the Department for liaison and cooperative arrangements with other agencies. Active participation of representatives from the Office of Management and Budget, the Department of Defense, and the Veterans Administration provides the Department with valuable information on the related experience and data requirements outside of the Department.

The U.S. National Committee on Vital and Health Statistics (USNCVHS)

A second committee, USNCVHS, established by the "Health Services Research, Health Statistics, and Medical Libraries Act of 1974, P.L. 93-353, provides a communication link between the public and private sector. This committee is the primary outside advisory group on matters of health statistics to the Secretary and the Assistant Secretary for Health. With staff support from the NCHS, the Committee stimulates studies and provides advice on statistical health problems of National or international interest, on health statistics and health information systems, and on a range of health statistical issues and activities that affect health care policies.

These two advisory bodies--the HDPC and the USNCVHS--provide a balanced and strong influence in the development of systematic statistical systems and mechanisms for assuring better coordination, integration, and accountability of these systems.

A Framework for Health Statistics Activities

P.L. 93-353 designates the NCHS as the focal agency for the collection and dissemination of the following data on the Health of the Nation:

(A) The extent and nature of illness and disability of the population of the United States, including life expectancy, the incidence of various acute and chronic illnesses, and infant and maternal morbidity and mortality,

(B) The impact of illness and disability of the population, on the economy of the United States, and on other aspects of the well-being of its population,

(C) Environmental, social, and other health hazards,

(D) Determinants of health,

(E) Health resources, including physicians, dentists, nurses, and other health professionals, and the supply of services by hospitals, extended care facilities, home health services, and other health institutions.

(F) Utilization of health care, including ambulatory health services and services of hospitals, and other institutions,

(G) Health care costs and financing, including the trends in health care prices and cost, the sources of payments and Federal, State and local governmental expenditures for health care services, and

(H) Family formation, growth, and dissolution.

In its wisdom, the Congress clearly recognized the need for the production of baseline, general purpose statistics to provide a broad range of data for the development and implementation of health care policy. It is also noteworthy that in the original National Health Survey Act of 1956, Congress recognized the need for quality data. In addition to specifying the kind of health data needed, Congress provided "For studying methods and survey techniques for securing such statistical information with a view toward their continuing improvement.

The Annual Report on the Nation's Health

P.L. 93-353 also specifies that the National Center for Health Statistics (NCHS) and the National Center for Health Services Research (NCHSR) assist the Secretary of DHEW to prepare and disseminate an annual report to the President and Congress. This annual report on the Nation's health entitled Health: United States, the first volume of which was issued in January, 1976, is an on-going mechanism for organizing the available data. Information from the various data collection systems in NCHS are combined with that from other components of the Federal Statistical System, including, for example, expenditure data from the Center for Disease Control, manpower statistics from the Bureau of Health Manpower, and the medical care price data from the Bureau of Labor Statistics. Non-Federal statistics such as those collected by the American Medical Association and the American Hospital Association were also included.

Work is proceeding on the second annual report which will summarize developments in the past year, and will analyze specific health issues and problems. Analytical chapters planned include such subjects as the health of children, hypertension, and current functioning and financing of the hospital industry.

Addressing Current Policy Issues

To illustrate further how the DHEW is responding to the data needs of policymakers in the health area, I will briefly discuss a few current illustrative issues and the emerging data requirements and implementation.

1. National Health Insurance - One of the major issues in the development of a health care policy is National health insurance. To provide Congress and the President with information about various options for financing health care, NCHS and NCHSR are collaborating on a Medical Care Expenditures Survey, that will begin in January 1977. It will produce National and regional data on utilization of and charges for medical care services, including hospitals, physicians and dentists, and on such expenses as drugs. Although National health care expenditures in the United States total about $125 billion in Fiscal Year 1976, little is known about the distribution of these expenditures, and what medical care services are used, who uses them, and who eventually pays the bill. The data from this survey will give information that relates to one of the major health policy issues facing the Nation. The information we are collecting and will analyze will help us make informed decisions about National health insurance and other major health issues. We will obtain vitally needed information about health insurance coverage, access to health services, and financing health care. The study will lay the groundwork for addressing and analyzing patterns of health services use and payment for future large scale health insurance programs.

2. Quality of Care - This has become one of the central health policy issues of the 1970's: The Professional Standards Review Organizations' program has been launched to develop a Nationwide system of hospital quality assurance. This quality assurance effort is expanding to include non-hospital institutional care and ambulatory care, and may well encompass care reimbursed by private health insurance as well as governmental programs. It has been recognized that the need for data is crucial to the successful implementation of the PSRO program. No issue better illustrates the current controversies besetting the program than the question of how data systems are to be developed, and who will have access to what type of information.

3. Health Planning - Up to this point, I have discussed health statistics at the National level. However, reliable health statistics at the State and local level are also necessary (Rice, Densen). Passage of the Health Planning and Development Act of 1974 emphasizes this need. As a result of this law, P.L. 93-641, the U.S. has been divided into 212 health service areas; within each of these, the law establishes a health systems agency among the functions of which are the compilation of available data pertaining to health status and its determinants and the utilization of health services. The law also specifies that the staff of each agency shall have expertise in the "gathering and analysis of data."

This mandate provides many challenges. For example, what is health and how does one measure health status? How do we measure utilization? Are there barriers to utilization and how do we identify them? What data are available that are suitable to assess these problems and how accurate are the data? Enactment of the Health Planning and Development Act of 1974 emphasizes the need for the statistician to provide timely and accurate data bases and skillful analysis and interpretation of the data at the State and local level.

A major resource to meet the needs of health systems agencies is the Cooperative Health Statistics System (CHSS). The CHSS is designed to provide the framework for a coalition among the various levels--National, State and Local, for the production of health statistics. Input is obtained from data users at these geopolitical levels regarding data they need, turnaround time they require, and the required frequency, reliability, etc. The agreed-upon data items are collected by the level best equipped to collect them and the data are then shared with all users. NCHS has the responsibility for providing National leadership by coordinating the efforts among Federal agencies and between the Federal government and the States, and in assuring the quality and comparability of data collected. The CHSS will help mold the current fragmented data collection activities throughout the country into a cohesive system. This should result in savings in data collection costs, a greatly reduced burden on data providers, and more important, data will be comparable and in the requisite detail for most uses by all participants in the system.

We in NCHS are convinced that the full potential of the Cooperative System will be reached most quickly and efficiently by the development of strong, well-staffed State Centers for Health Statistics in each State. These centers would serve as focal points for the assembly and analysis of all health data and have responsibility for primary data collection and activities for at least one of the statistical components of the System. The centers should also have the capacity for analysis of health-related data, from whatever source, and be able to provide consultation to planning agencies and other users on the limitations and potentials of available data as well as on needs for additional information and how it can be secured. Several states now have such centers, and we are hopeful that others will follow their lead.

The Statistician's Role

But can one talk about the role of statistics in developing policy, whether in health or in any other field, without talking about the responsibilities of the statistician? Statistical information is generated to be used; without a statistical interpreter to explain such characteristics as reliability, precision and accuracy, data are weak, sometimes misleading and easily misused. Borrowing from Sir Claus Moser's speech at last year's ASA meeting the "...Foremost responsibility (of the statistician) is to contribute to more enlightened and efficient 'decisionmaking' through the provision of a timely and accurate data base and through the fullest possible exploitation of our skills in analyzing and interpreting the data." This responsibility is key whether we are government statisticians at either the Federal, State or Local level, or working in academia or in private industry.

The statistician's responsibility whether employed in the public or private sector, is to two audiences: 1.) the users of his product and 2.) the suppliers of his raw material (Frankel). In discussing the development of policy, the focus is on the use of statistical information. Data may be used in health decisions concerning the level of funding of community mental health centers, of the new health systems agencies, and of medical facilities. Data may also indicate tradeoffs for resource allocation.

The statistical environment appears to be shifting. Until recently, the health statisticians have viewed their responsibility primarily as one of providing accurate data. However, with the increasing demand for hard data to assist in the decision-making process, it is no longer sufficient to provide accurate statistics. It is necessary to develop on this foundation of technical expertise a framework which expedites the provision of information. Thus, we need to consider new ways to provide data in a timely fashion and in an analyzed, packaged format. Among the various ways of providing timely data are data sharing, piggy-backing surveys, and anticipating user needs.

First, let me point out that with unlimited resources, i.e., money and manpower, getting statistics out in a timely fashion would be much less of a problem than it is now. However, the statistician rarely operates with unlimited resources. We are searching for innovations which enhance data production with minimal increased demand on resources. I have mentioned three ways--data sharing, piggy-backing, and anticipating needs--let us examine more thoroughly and point out specific action in health data programs:

1. Data Sharing - As data become more costly to collect and our concern for respondent burden, alternative methods for obtaining the necessary information must be actively identified (Oaxaca, Dunn). The Health Data Policy Committee is working towards standardization of terminology and analytic conventions. While this will facilitate the multiple use of various data sets, the privacy and confidentiality of the respondent must be respected.

Data sharing involves not only collection but also the dissemination of such data. One such mechanism is through the timely development, distribution, and sale of data tapes.

2. "Piggy-backing" Surveys - By adding a few key questions to an on-going survey, time and costly implementation procedures can be spared. The forthcoming swine flu supplement to the Health Interview Survey is an example. Beginning in September, questions will be asked as to whether or not the respondent received the vaccine; if so, where, and if not, why not?

3. Anticipation of user needs - What is required is the foresight to put into place statistical systems that will be producing reliable and current information by the time the users become aware of their need for such data. We need statisticians who combine knowledge about our social and economic institutional structure with a high level of methodological sophistication. Such statisticians are at present in short supply. Certain alterations in the curriculum of our universities may be necessary if the demand for such statisticians is to be met.

In addition, Federal health statisticians are becoming increasingly aware of the uses and abuses of their products through communication channels of advisory groups such as the Health Data Policy Committee and the U.S. National Committee on Vital and Health Statistics. Experience has shown that various data collection agencies and users can cooperate to resolve issues in the provision of timely and relevant data.

We have identified three approaches to the provision of analyzed and packaged data:

1. New and Improved Methodology. As exemplified by the Medical Care Expenditure Survey, new surveys are time consuming to design and costly to implement. While new surveys may be the only way to answer some questions, they may not be the universal solution. More imaginative use can be made of existing data; one such NCHS project is to develop mortality indicators suitable for small area analysis.

2. Packaging of Information. Data needs to be presented so that it becomes information rather than simply a collection of numbers. More extensive use should be made of media such as chart books and tabular compilations such as those presented in the Annual Report to Congress. The development of computerized graphics is also advancing. Software which provides sophisticated, multi-colored graphs is now available. NCHS is using one of these packages to produce a National Atlas of Mortality.

3. Statistical Training. In recognition of the importance of the communication between the statistician and the users of statistical information, in-depth programs, where students of various disciplines work together in a problem-solving environment are needed. Several universities (University of North Carolina, University of Arizona, and University of Hawaii) have established these types of degree programs. The NCHS also provides short-term training through the Applied Statistics Training Institute (ASTI) program.

With this type of training, statisticians become members of the policy formulating team. Involvement in the planning phases, whether in legislating new programs, setting up clinical experiments, or designing survey collection will prevent: 1.) some of the goal inconsistencies pointed out by the recent legislation; 2.) Casually or haphazardly collected data; and, 3.) missing data such as to encourage biases (Finney).

For the next few years, the statistician's prime responsibility will continue to be the provision of timely and accurate data bases and the interpretation of this data. However, user needs will grow in keeping with changing societal programs; even the characteristics of the audiences may change. So too, will the statistical environment change with the introduction of new resources and technological advances. Thus, the statistician must continually be aware of his audience and the environment in the provision of information for the development of public policy.

Privacy Protection

I have only alluded to the problems of confidentiality in my discussion thus far. I cannot end without expressing my concerns in this important area that is seriously affecting the statistical environment.

After years of struggling with the issues around confidentiality, NCHS--like other Federal government agencies--last year went to work to implement the Privacy Act of 1974. Actually, we had already for many years operated under a Section of the Public Health Service Act (Section 308 (d)) which precludes our using information for any purposes other than those for which it was collected and prevents our releasing information about individuals or establishments without their permission. While there were new and specific requirements in the Privacy Act which we had to meet, they gave us no difficulty.

Nevertheless, there remain some serious issues and concerns in the confidentiality area, including these:

1. The needs of our sister research and statistical agencies for legislation similar to ours, permitting them, too, to offer strong confidentiality protections (Duncan).

2. There are pressures for the passage of additional, stronger, and broader laws giving privacy protection, and there are, no doubt, some needs for them. But the laws must be written very carefully to assure that they do not unnecessarily hamper research.

3. More attention must be given to the problem of unintentional disclosures of information through statistical publications and through the release of public use micro-data tapes. Here, too, the rules must be written so that they do not unnecessarily impede statistical research.

4. As States begin to pass their own privacy laws, some of them very stringent, there is a pressing need to develop and promote model State laws which will provide needed protections but at the same time permit the efficient use of statistics to meet all the real needs for them at local, State, regional, and National levels.

The basic conflict between the citizen's right to privacy and society's need to know may never be resolved to everyone's satisfaction, but that is no reason not to keep trying. Our responsibility is to those who supply our raw material as well as to those who use our data.

In conclusion, I want to emphasize again that statistics and statisticians are going to play an increasingly important role in the development of health care policies. We are being asked to produce more data and more relevant data with resources that are not growing commensurately. We are being asked to aid in the interpretation and analysis of the data as well. It is an exciting prospect and one that will tax all our expertise and ingenuity.

REFERENCES

1. Abel-Smith, B., Value for Money in Health Services, Social Security Bulletin 37(7): 17-28, 1974

2. Abelson, Philip H., Editorial: More Laws, More Complexity, Science 192(4246): 1291, 1976

3. Andreano, Ralph L., Wiesbrod, B.A., American Health Policy, Chicago: Rand McNally College Publishing Company, 1974

4. Austin, Charles J., The Politics of National Health Insurance: An Interdisciplinary Research Study, San Antonio, Tx:Trinity University Press, 1975

5. Ball, R.M., Response to Medical Care Guarantees: Economics of Choice in, Implications of Guaranteeing Medical Care, J.G. Perpich (Editor), Washington, D.C.: National Academy of Sciences, 1975

6. Cooper, T., Do We Have a National Health Policy?, Journal of Occupational Medicine 17(10):625-627, 1975

7. Densen, Paul, Presentation at Public Health Conference on Records and Statistics: St. Louis, Missouri, June 1976

8. Duncan, J., Confidentiality and the Future of the United States Statistical System, American Statistician 30(2):54-59, 1976

9. Dunn, E., Social Information Processing and Statistical Systems--Change and Reform, New York: Wiley, 1974

10. Finney, D.J., Problems, Data, and Inference, Journal of the Royal Statistical Society A 137(Part1): 1-19, 1974

11. Frankel, L., "Statistics and People - The Statistician's Responsibilities" Journal of the American Statistical Association 71(353): 9-16, 1976

12. Fuchs, V., Who Shall Live?, New York: Basic Books, 1974

13. Hauser, P., Social Statistics in Use, New York: Russell Sage, 1975

14. Illich, I., Medical Nemesis, New York: Pantheon Books, 1976

15. Kahn, Herman, The Next 200 Years, New York: William Morrow and Co., 1976

16. Kendall, M., Statisticians--Production and Consumption, American Statistician 30(2): 49-53, 1976

17. Merriam, I.C., Social Security Policy Formulation and Health Research, International Journal of Health Services 3(1):59-68, 1973

18. Mueller, M.S., Gibson, R.M., National Health Expenditure, Fiscal Year 1975, Social Security Bulletin, 39(2):3-21, 1976

19. Murnaghan, J., Health Services Information Systems in the United States Today, New England Journal of Medicine 290(11):603-610, 1974

20. Oaxaca, F., President's Reporting Reduction Program, Statistical Reporter 76-10:185-190, 1976

21. Public Health Service, Forward Plan for Health, Washington, D.C.: DHEW, 1975

22. Public Health Service, Health Statistics Plan, Washington, D.C.: DHEW, 1975

23. Rice, D., Presentation at Public Health Conference on Records and Statistics, St. Louis, Missouri, June 1976

24. Rice, D.; Wilson, D., The American Medical Economy--Problems and Perspectives, forthcoming in Journal of Health Politics, Policy and Law

25. Russell, L., Inflation and the Federal Role in Health, in, Health: A Victim or Cause of Inflation?, New York: Prodist, 1976

26. Rutstein, D., Berenberg, W.; Chalmers, T.C.; Child, C.G.; Fishman, A.P., Perrin, E.B., Measuring the Quality of Medical Care: A Clinical Method, New England Journal of Medicine 294(ii):582-588, 1976

27. Saward, E.W., Medicare, Medical Practice, and the Medical Profession, Public Health Reports 91(4) 1:317-321, 1976

28. Tukey, J., Methodology, and the Statistician's Responsibility for Both Accuracy and Relevance, Statistical Reporter 76-13:253-262, 1976

29. Wennberg, J., National Health Planning Goals, in National Health Guidelines Volume III., Rockville, Maryland: Health Resources Administration, 1976

30. White Paper: Towards a Comprehensive Health Policy for the 1970's, Washington, D.C.: Department of Health Education, and Welfare, 1971

31. White, R.L., The Right to Health: The Evolution of an Idea, Iowa City, Iowa: University of Iowa, 1971

DESIGN AND METHODOLOGY OF THE SOCIAL SECURITY ADMINISTRATION'S EVALUATION AND MEASUREMENT SYSTEM

Jacob Deutch, Social Security Administration, HEW

In a review in this organization's journal of Eugene I. Grant's text, Statistical Quality Control (4th edition coauthored with Richard S. Leavenworth), W. E. Deming praises it as the outstanding text for both teaching and home study of the statistical control of quality. Deming does find in it, and all other such texts, however, the limitation of treatment of the controls of the process to the assignable variation or defect that the worker himself can correct. The main cause of these variations and other production troubles, says Deming, lies mostly in the system itself. The production worker works within the system; i.e., policies or procedures. He is not responsible for it, nor can he change it. Only management can change the system. This is the forgotten half or forgotten 80 percent of quality control, says Deming.

Expanding on this point of view in the August 1975 issue of Interfaces, a journal of the Institute of Management Sciences, and drawing upon his renowned efforts and experiences in quality control and in methods of administration in Japan, Deming now raises to 85 percent the common or environmental faults that stay in the system until reduced by management. The cause specific to the production worker or machine are now seen to be a mere 15 percent.

Paraphrasing Deming, a major roadblock to quality in America and a guarantee of future disappointment is the supposition all too prevalent that quality control is something that you install like a new Dean or Commissioner (or new carpet or new furniture). Install it and you have it. Actually, quality control, to be successful in any company, must be a continuous learning process, probably more at the top than at the bottom and under competent tutelege.

The second roadblock sighted and cited is management's supposition that the production workers are responsible for all the trouble; that there would be no problems in production or in service if only the workers would do their jobs the way that they were taught. Nirvana!

It is Deming's experience that it is something new and incomprehensible to a man in an executive position that management could be at fault in the production end. Production and quality in the view of management are the responsibilities of the production worker. Research into faults of the system, to be corrected by management, is not what a manager is trained for. The results are the retention of the faulty system, high costs, low quality, and persistent quandry why efforts at amelioration are to no avail.

Management reacts to this situation by turning the job over to an office of quality control. This would be a happy solution if it solved anything. It usually doesn't because what we gain are new titles and perhaps higher grades or new or old people with tried and tired ideas. A recipe that guarantees failure.

Statements by management of aims desired in quality and production are not quality control, nor are exhortations, pleas, and platitudes addressed to the rank and file very effective. The results are merely intensification of guerilla sniping. Something more is required.

It may have occurred to some of you that if I substituted case worker, interviewer or adjudicator for the term production worker and if I had added legislators and public to management, I would have been describing the situation in the administration of welfare benefits today. The Social Security Administration's Evaluation and Measurement System is the something more. How much more is to be determined.

In the early 1960's, a team from the General Accounting Office studied how policies and procedures (i.e., the system) were established in the Social Security Administration for handling claims for retirement and survivors insurance benefits and how these policies were observed in practice. Their report to Congress highlighted two conclusions. One, SSA's policies and procedures were too lax, leading to incorrect payment of benefits, and two, SSA had no continual and objective means of determining the validity of its policies and procedures. This report served as the final impetus which led to the establishment of SSA's Evaluation and Measurement System (EMS). This system in turn, has rated GAO's conclusions, and has given it a grade of 66.7 percent. A description of the rater, that is EMS, is the heart of this paper. A justification of the rating will also follow.

The Commissioner, subject to the same type of attack when he for a brief period headed up what is now the Social and Rehabilitative Services Administration, immediately agreed to set up a continuous and objective measurement system but with one exception refused to modify the existing procedures specifically attacked in the GAO report pending results from the new EMS system. Among the procedures under attack were those establishing date of birth, proof of marriage, identity, whether the young wife or widow has a child of the wage earner in her care, proof of support, and retention of documents submitted as evidence. In essence, the heart of the claims adjudication process was attacked.

While there was no continuous and objective means for determining validity of policies and procedures, the product or service was of course not designed and evaluated in a vacuum. There were frequent ad hoc studies, feedback from operating personnel, claimants, and of course, congressmen. It would be pollyanish to presume that the feedback was unbiased but the ad hoc studies were attempted in good faith. Nevertheless, these studies (as well as many quality appraisal efforts) were conducted within operational contexts and constraints resulting in modifications of probability samples (not in its selection but in differential completion rates) thus destroying not merely the possibility of measuring precision but most importantly the representativeness of the results, and with insufficient controls and standards of measurement under which doubtful creditability could be attributed to results no matter how small the sampling error.

To help design an objective system Joseph Steinberg was lured from the Bureau of the Census. He received the assistance of representatives, managerial and operating personnel of the various bureaus within SSA in designing the system. Draft after draft were critically reviewed, attacked, and discarded. Finally, with some arm twisting and tongue lashing from the Commissioner a design acceptable to all was adopted. The period of gestation approached 9 months. Some things can't be hurried. A point to be stressed is that a system which will more likely gain the acceptance of the rater and the rated is one that is jointly produced even though in the course of joint production the legitimacy of the birth of those involved are reflected upon by each other. Happily for me I did have a certified copy of my birth record readily available and a furtive glance at it put my fears to rest.

Some of the alternatives faced and decisions made in designing this system will be referred to in the remainder of this paper. One decision that could be made only by the Commissioner was where EMS should be placed organizationally. In order to maintain credibility, it was decided that it could not be located within those arms of SSA responsible for policy and procedure formulation or implementation. EMS was placed in the Office of Research and Statistics (ORS). Under a recent reorganization EMS was culled from ORS and both organizations report directly to the Associate Commissioner for Program Policy and Planning. While there are many reasons for an overall quality control or quality appraisal, quality assurance or evaluation (the same concept may masquerade under different nomenclatures) not to be lodged under those immediately responsible for designing and/or performing the operation or service, we must be aware that while we are avoiding the Scylla of seduction by operational needs or interests we are tempting the Charybdis of being entirely ignored. This is a familiar problem facing each maiden including one whose name is Quality Appraisal or Evaluation unless closely chaperoned by the very top executives.

Now back to the design and operation of EMS. Each month a stratified nonself-weighting probability sample of 1,000 claims for retirement and survivors and disability (and recently supplementary security income) benefits that were processed, awarded or disallowed benefits the prior month are selected by computer. Stratification is by factor of eligibility to benefits and is frequently modified as greater errors are indicated or more detailed information is required regarding selected factors of eligibility. All claims material pertaining to the case residing in our district or reviewing offices and wage record data in Baltimore are transferred to EMS for reexamination by one of our policy measurement specialists. These specialists have usually had district or reviewing office experience as well as additional training in statistics. Data regarding the policies or procedures applied in the original claims handling, the evidence submitted, the findings made, the apparent correctness or deficiencies of the findings under existing operating standards, demographic characteristics of the claimant, operational data, amount of benefits involved, etc., are entered into a relevant subset of 18 optical scanning coding sheets. The punchcards subsequently produced are fed into the computer for editing of completeness and for internal logic. When the data pass these tests, they become part of our data base. This data base presents an early detailed source of information to decision makers regarding the existent claims process and can frequently answer the question, what if...?

In the meantime a transmittal sheet containing the name and address of the selected claimant as well as the names or initials of the district office employees previously involved in the case are forwarded to the servicing district office together with relevant questionnaires. Upon receipt of these forms the district office manager, in our attempt to reduce bias that may result due to knowledge of what has already occurred in the case, withdraws any material regarding the case from the district office's open files. He then assigns the case to a nonpreviously involved district office management or technical employee in accordance with a previously designed sample scheme. In designing the system, preference was to have the field redevelopment work done by an outside organization or by non-DO based SSA employees but cost considerations militated against these approaches. Hence the substantial efforts to mitigate any bias effects by the district office data gathering function in the system.

The assigned district office redeveloper writes to the selected beneficiary that his case has been selected by chance in order to evaluate our methods and procedures, and not because we believe there was anything wrong with the prior decision, and that he will be phoning him within a few days to arrange an appointment at the beneficiary's home.

At the beneficiary's home, the district office redeveloper begins the interview by reading the following statement: "Every month we choose a small number of new cases for review. The purpose of this review is to help us evaluate our policies and procedures. Your case was selected purely by chance and is part of a statistical sample. Although the main purpose of the interview is to judge how well our application process is working, if we should find an error we will correct your benefit amount. Completion of this questionnaire is not compulsory. In any event the Social Security Act requires verification of eligibility factors and it will be necessary for us to contact independent sources. The statements you make or information you give us will be used only for statistical purposes and to determine your correct benefit amount. First, I would like to ask you a few general questions." He then proceeds to ask questions from the biographical questionnaire and various supplements designed to elicit information regarding the claimant's birth and baptism, previous residences school attendance, marital history, military history, employment history, etc., as they pertain to the possible establishment and existence of (documentary) evidence regarding the factors of eligibility at issue. The questions are to be asked in a structured manner and initially, precisely as worded.

This questionnaire merits further attention. First it should be noted that its chronology is cylical. We trace residence chronologically, then schooling, then marriage, children, employment, etc. Second it should also be noted that many questions obtain the desired information indirectly. Thus responses to an individual's age come not only from asking when you were born but where and when did you enter school. Thus if you first entered school in City X and you moved from City X in 1903 then based upon the date of birth of 1901 that you originally gave us you were only age 2 when you entered school. I am talking to a child prodigy! No! Let's start again! The questionnaire thus permits additional reinforcement or weakening of the allegations with every additional question. This pattern of questioning is used to considerable advantage in obtaining more accurate information, than in the regular claims process, regarding income, assets and living arrangements in the Supplementary Security Income Program. In addition to questions regarding the claimant's birth, citizenship and marital status, information is also obtained regarding the claimant's children and their ages and residences. The claimant's residences and living arrangements over the past 3 years rather than merely the current residence and living arrangements are pursued. Current household and personal expense questions are asked in addition to questions regarding income and resources. The questionnaire webs a net around the allegations either strengthening or weakening the allegations with every additional question. This pattern of questioning is used to considerable advantage in obtaining information regarding income, assets, and living arrangements in the Supplemental Security Income Program.

Upon the conclusion of the questionnaire, apparent inconsistencies or doubts raised by looking at the information supplied, either the specific allegations, or the allegations in their entirety, call for additional probing. After reconciliation or modification of the allegations, the sample person is then requested to sign a consent statement to permit SSA to obtain the required evidence from the various custodians of the records.

Subsequently, contact with the custodians of records and other third parties are made to obtain all the evidence or explanations necessary to establish all the factors applicable to the case. These include identity, date of birth, current and prior marital relationship, full-time school attendance, child relationship, child dependency, parent-child relationship, in-her-care, date of death, lump-sum death payments and more recently with the advent of the Supplemental Security Income Program, income, resources, and living arrangements.

Several recontacts with the claimant and others may be necessary to reconcile discrepancies among the allegations and evidence. The level of evidence (as well as the reinterviews) required to reach a finding as close to the truth, as feasible within the resources available to the EMS system, far exceeds that required in the regular adjudication of the claim.

Upon completion of the redevelopment (i.e., reinterviews with claimants, new interviews with third parties, record checks, and reconciliations), the completed forms and evidence are sent by the district office to the EMS staff, and this subsequent information is placed in a new folder and analyzed independently by a policy measurement specialist other than the one who had reexamined the case initially. If the redevelopment meets the EMS standards it is coded and fed into the computer. If not, it is returned to the district office for further development or reconciliation. The computer edits the data of the redevelopment for completeness and logic and when these tests are met the data are entered into the data base. The computer then compares this new information with the original information in the data base. All contradictions between the two parts of the case are printed out by the computer. It is at this point that the two separate folders are brought together and analyzed by a third policy measurement specialist in order to determine:

1. whether a true rather than a clerical coding difference exists;
2. what is the correct finding based upon all evidence in both files;
3. the substantive nature of the difference;
4. the reason for the difference;
5. the effect of the difference, and
6. the money amounts involved.

The above conclusions are entered into the data base. In addition a narrative summary of the case is prepared including how the difference was uncovered and alternative corrective management actions required to minimize such differences.

All cases involving differences or issues are made available to policy officials immediately for their information or comment. Subsequently all difference cases and a subsample of the nondifference cases are returned to the regular claims process reviewing office, which reviews the case including all the new material and informs EMS what it believes the true findings to be, and proceeds to any revisions in the prior decision, or sends a closeout thank you letter to the claimant. It also sends a closeout letter in the remaining nondifference cases. If the reviewing office reaches a conclusion different from EMS, it is reviewed in EMS for possible revision of EMS's prior decision. EMS's decision is purely a statistical decision. It is the reviewing office's decision that has the impact on the beneficiary.

When 6 months samples are 95 percent complete, a formal preliminary report on these 6 months as well as the cumulative 5-year period, is forwarded to the Commissioner, members of his executive staff, and interested policy officials.

When a 1-year sample is 99 percent complete, a much more detailed final report is prepared. Special reports and memoranda, formal and informal and oral presentations are made to any or all of the above when considered meaningful, either in the course of policy formulation or subsequent evaluation.

Now, let me return to the 66.7 percent rating that EMS has given the GAO report. A full 50 percent is immediately given to one of the two conclusions, viz, the need for a new continuous objective measurement system of SSA policies and procedures. The immediate acceptance and implementation of such a system substantiates this critique. But what about the other critique -- that of laxness of policies. Yes, EMS has established that some policies were too lax. But on any reasonable cost benefit analysis, EMS has also established that some were in the right ball park, and in the case of others, substantially less effort on the part of the Administration or the claimant would yield comparable results. Thus this critique is only one-third correct which when added to the full 50 percent grade on the other critique explains the mystery of the 66 1/3 grade.

While one of EMS's major tasks is to report to the Commissioner on the current state of claims policies and procedures, another major task is continual liaison with policy officials in policy establishment or modification to determine what data are necessary in policy formulation or in a choice of one policy among several alternatives and what data are required in subsequent evaluation thereof. While we have made significant progress in this latter function, and perhaps more progress than could be expected of an evaluation function, we are nevertheless traversing a rocky road that needs constant repair. Need I state that the objectives, constraints, time perspectives and value judgments among those in operations are not necessarily the same as those in policy or systems design or quality review. But nevertheless, success breeds success and the significant assistance that EMS has been able to provide policy and operating officials in such areas as establishing date of birth, a problem that eluded successful resolution for 25 years, how to deal with complex legal marital and child relationship issues, whether to recall material purged from inactive files upon subsequent actions in the case, upon the proper grade level of employees to certify evidence, are some examples which have established a secure and growing role to the EMS process. Operating in a research setting permits more independent pursuit of meaningful paths of exploration and analysis. On the other hand there is risk of oblivion, as those on the firing line ignore an evaluation body's existence or ignore its results. Careful attention to the needs of operation and policy officials and demonstration of the applicability of the research efforts in decision making, and the persistent but sensitive conveyance of these findings to the proper officials without any intent or appearance of an advocacy or adversary role is the only feasible approach. Some of the improvements in the claims process flowing from EMS were a result of highlighting the need for training but most meaningful improvements called for changes in the procedures or systems involved.

The integral role of the system is now widely and deeply ingrained in the SSA claims process and it has recently been called upon to play a critical role in our new Supplementary Security Income Program. The EMS task, much more than the regular claims process, involves sophisticated and intense but extremely delicate and sensitive probing of an applicant's income, resources and living arrangements. We approach this assignment with due modesty as this task according to the headlines of your daily newspaper has not been successfully accomplished to date at any level of government or elsewhere.

Insufficient time has elapsed since incorporating a subsample of SSI cases for me to be able to relay any meaningful results. Suffice it to say that we have adopted the same approach in these cases as in the case of retirement survivors and disability insurance (RSDI) claimants. As indicated previously this approach includes examining all prior material available in the case such as prior claims for RSDI benefits, all wage records available to SSA, a structured questionnaire not merely asking directly the person's income, resources and living arrangements but also his expenditure patterns, residences, number and location of children, etc., as well as subsequent confirmation of allegations with custodians of records and third parties. From this web of interweaving allegations we believe that we will gain a better vantage point from which to determine not merely the validity of the claimant's allegation but be able to offer meaningful conclusions regarding proper alternatives to the current intake and redetermination procedures.

THE APPLICATION OF QUALITY CONTROL METHODOLOGY IN THE AFDC PROGRAM

Sue Ossman, Social and Rehabilitation Service, DHEW

Introduction

Before discussing the application of quality control methodology in the AFDC program today, let me briefly tell you how the AFDC program operates and why the quality control system was developed by the Department of Health, Education and Welfare.

Aid to Families with Dependent Children is a program administered by the States who set the amount of payment for each needy family. All income and resources of the family must be considered by the State when determining size of the payment. The Federal government, through the Department of Health, Education and Welfare, sets program guidelines for the States and reimburses them for somewhere between 50 percent and 83 percent of the total program cost, depending on each State's per capita income. More than $800 million a month, or over $9 billion a year, now go to help 3½ million families consisting of 11½ million children and adults.

Three major factors impinge on the proper and efficient administration of the AFDC program. The first factor is State and Federal law and policy. These specify what families are eligible to receive financial assistance and the requirements regarding the administration of the program. They determine whether a person qualifies for an eligibility group--which needs are to be met and by how much money; which financial resources are to be considered and how they are to be measured in the payment determination; how to identify changes in circumstances which affect eligibility and the amount of the payment; and how all such information is to be submitted to and processed by the agency.

The second factor, is agency staff, skill and accuracy. These play a large role in carrying out the assigned tasks of the agency. Involved here are such elements as understanding manuals and instructions, previous training and work experience, size of individual caseloads, distribution of staff within operating units, and organizational structure within the local agency.

The third factor, and not the least of them, is client/agency administration in providing the agency with accurate and timely information. This includes knowledge of requirements, willingness to provide information, and ability to contact the agency about changing circumstances.

To ensure that these factors produce an eligible caseload that receives the correct amount of money, HEW has developed a quality control system which is operated by the States.

Objective of Quality Control

The quality control system in AFDC is an administrative program designed to determine the extent to which those receiving assistance payments are eligible and, if eligible, the extent to which they are receiving amounts to which they are entitled. Not only is the objective to measure the extent of ineligibility and incorrect payment, but also to hold the incidence of error below pre-established tolerance limits. It accomplishes these objectives by means of: (1) continuous review of valid and reliable cross-sectional Statewide random samples of AFDC cases in all States, District of Columbia, Guam, Puerto Rico and the Virgin Islands; (2) periodic assembly and analyses of case findings to determine incidence, amount and reason for occurrence of errors; and (3) application of corrective action to reduce the causes of excessive error rates.

To ensure that State QC systems are operating in accordance with Federally established requirements and to assist each State agency in fully utilizing its QC system, Federal staff conducts ongoing appraisals of State QC operations. The Federal appraisal consists of: (1) annual comprehensive reviews of all eight components of the system and (2) ongoing re-reviews of a subsample of State QC reviews. For the Federal agency, the appraisals serve two purposes: (1) to confirm or deny the validity of State error rates; and (2) help to pinpoint problems for which Federal technical assistance can be provided to any component of the system.

I shall first discuss the components of the State QC system, providing somewhat more detail on the components of most interest to survey researchers. Next, I shall describe briefly our Federal monitoring system and share with you a few of the operational problems that we have encountered in our survey research methodology.

Components of the State Quality Control System

The eight components of the State QC system are so interrelated that they constitute a continuous cycle starting with the sample selection process. This is the mechanical step of selecting the sample cases from the total State caseload. QC sample sizes are based on the assumption that an ineligibility case rate of 3% is reasonable for States to achieve (although there is some difference of opinion on the validity of this assumption at the present time). Sample sizes are, therefore, based on an ineligibility rate of 3% with a precision of 3% for States with relatively small caseloads up to a precision of 1% for States with relatively large caseloads. This translates into State sample sizes ranging from 150 cases up through 1200 cases--about 45,000 cases nationally.

Systematic random sampling is applied to a sample frame of active cases receiving a money payment in each month. (Since the review period covers six months, one-sixth of the sample cases are selected each month.) Usually the frame consists of the actual payroll listing for the month. Cases which are not included in the QC review system are generally eliminated after selection of the sample. Oversampling is required to compensate for this.

In sampling of this type, it is important for the structure of the frame not to be cyclical in order to avoid biased results. This is monitored carefully by the Federal regional offices. Any changes in sample frame structure or State sample selection techniques cannot be made without prior approval of the Federal regional office.

The second component is the review process which produces the raw data on which the activities in the later components are based. This process requires meticulous attention to detail, a firm knowledge of the welfare regulations applying to the cases reviewed, expertise in the principles and techniques of interviewing and investigation, and sound judgment for drawing conclusions. This component is primarily composed of two parts. One part is the analysis of the case record; a second part consists of verifying elements of eligibility and payment for the month of review by a full de novo field investigation. This investigation includes contact with the client and contact with collateral sources of information. The process also includes correspondence, review of documents, telephone conversations, computation of a budget by agency standards of assistance and any other activity pertinent to the review of the case.

The case record is the repository of all current information about the case upon which eligibility, amount of need, and amount of payment is based. Through analysis of the case record, the reviewer familiarizes himself with the family situation, notes the specific facts related to conditions of eligibility and payment, and identifies gaps or deficiencies in information. Where documents or statements are contained in the case record, the reviewer identifies those which may be used as verification. All relevant information obtained from the case record analyses is recorded on worksheets. All documentation information must include such specific information as volume and page reference to public records.

Having analyzed the case record, the reviewer is now ready to start the field investigation. The personal interview is probably the most important part of the case review since the client furnishes most of the evidence necessary to establish his eligibility. If the client does not have the necessary evidence, he often advises the reviewer where the required verification can be obtained.

The reviewer structures the interview to ensure adequate coverage of all eligibility and payment factors. When the case record does not contain adequate verification of an eligibility factor, the reviewer must obtain additional verification. For example, information on military service and work history means there is a possibility of veteran's benefits and company pension rights. These factors are discussed with the client to establish the possible availability of these benefits.

Pre-planning and structure are essential to assure complete coverage of all eligibility factors. However, the QC interview, per se, is not structured or directed in such a manner as to preclude the possibility of the client's active participation. Relevant topics are covered in such a fashion as to permit the client freedom to discuss his situation. The reviewer furnishes the lead topic for discussion, giving the client the opportunity to explain past and present circumstances freely. The circumstances discussed, together with the case record, gives the reviewer leads for the second part of the field investigation, which is collateral contacts.

The QC review requires independent verification of all applicable elements of eligibility, unless the case record contains documentation of all necessary verification. When a client makes a negative response, the reviewer must build a solid basis for deciding that the client, in fact, does not have resources available. He must also check out resources declared by the client to determine if the client was correct. Such clearances as the Social Security Administration, Internal Revenue Service, Employment Security, etc., are almost routine.

When all the necessary information has been obtained, the reviewer analyzes the documents and verifications in terms of the case situation as of the month under review. Every effort is made to reach a definitive conclusion on each sample case with respect to eligibility and correctness of payment according to the State plan. Once a definitive conclusion is reached, the reviewer summarizes the case findings on the review schedule according to instructions provided by the Federal agency. (As cases with incorrect payment amounts are found, they are referred to the local agency for individual corrective action. This, however, is not the ultimate goal of the QC system. The ultimate goal is to identify the causes of the repetitive type of errors and to eliminate them through corrective action.)

The third component of the system is data management. Here, the completed schedules are edited, tallied, and compiled into statistical reports which States are required to submit to the Federal agency covering the results of the QC reviews in each reporting period. These reports include sample size; number of ineligible, overpaid and underpaid cases and payments; distribution of error cases by responsibility for error (i.e., agency or client) and by primary program element in error; and cross-tabulations of various case record characteristics by error and non-error cases.

Taken as a group, these reports, show QC error findings from various perspectives for purposes of analyzing probable causes of error. The basic QC case schedules contain additional information which many States use for their own specialized needs.

The fourth component is data analysis. To be an effective guide to corrective action, data analysis should be a continuous process involving information of major concentrations of error both in number of cases and in payment amounts; evaluation of previously implemented corrective actions; and, to the extent possible, an examination of the cost-benefit implications of recommended actions.

The data analysis process breaks down into two main areas--one is the development of profiles of error cases and the other is the development of profiles of error-prone cases. The primary difference between the two is the population referred to. Profiles of error cases refer to percentages, or probabilities, of error when only error cases are considered. Profiles of error-prone cases refer to percentages, or probabilities, of error associated with a review of cases in the total caseload.

Data analyses may point out some or all of the following: which sources are causing most of the errors; which of the errors are most important; the locations and conditions which need closer study and scrutiny; and numerous other points relevant to the control of quality.

The fifth component, distribution of QC findings, provides for information at appropriate intervals to agency staff--to top administrative staff for corrective action; to training staff for intensified training in selected areas; to eligibility workers for identification of errors; and to quality control staff for information on the total case review findings.

The last three components of the system make up the three corrective action phases -- program analysis, corrective action planning and corrective action implementation. In most, if not all, of the State public welfare agencies, these three components are handled by a corrective action panel consisting of top executive staff representing the various disciplines and expertise of the agency. Administrative improvement changes resulting from corrective actions taken may affect Federal and State policies, eligibility determination procedures, agency performance, and client reporting. The cost effectiveness of the corrective actions taken and the effect of changes on caseload error rates are evaluated in subsequent sampling periods.

Federal Re-Review

Each State agency is required to submit to the Federal regional office a listing of all sample cases it has selected for its quality control review. As case reviews are disposed of, photocopies of the case schedules are made and submitted to the regional office. It is from these photocopies that the Federal re-review subsample is selected, also by the systematic random selection method. All photocopies are checked against the State sample list to verify that the disposed of case was part of the State's original sample. (Incidentally, initial State findings, including the amount of payment and the amount of error, are recorded for each case at the time the Federal subsample is selected. This is important in the determination of final error rates, which we will point out later.)

Federal subsample sizes range from 70 cases, or roughly 1 out of every 2, in States with the smallest QC samples to 180 cases, or about 1 out of every 7, in States with the largest QC samples--a national total of approximately 8,000 cases. If the sampling interval is small, random digits are selected within each interval to ensure that the selection of subsample cases is, in fact, random.

The State quality control file as well as the case record on each of the selected re-review cases are carefully examined, checking off areas requiring a field review for additional information. After obtaining the findings of the field reviews (or at the conclusion of the examination of State files), the regional office reviews the eligibility of the re-review case and recalculates the amount of payment. If the findings of the Federal re-review differs from the State quality control results, a formal meeting takes place between Federal and State staff to determine the correct findings of eligibility and amount of payment to the case. The conference discussions may lead either to corrections in the re-review findings, corrections to State review findings, or to an unresolved difference. Unresolved differences are referred to the Regional Commissioner for final decision or resolution. If the results of the discussion reverses the Federal re-review findings, then no difference between State and Federal results is recorded; if the Federal re-review findings are sustained, then a difference is recorded.

Double Sampling Regression Methodology for Determining Final Error Rates

A Federal re-review sampling essentially constitutes a "double sample". A double sampling regression methodology was developed for the Social and Rehabilitation Service by Mr. Morris H. Hansen of Westat Research, Inc. It involves finding the relationship between final Federal findings and State original findings in the Federal subsample for determining case error rates. This relationship may range in value from poor to perfect. The value is calculated separately for ineligible cases, eligible but overpaid cases, and eligible but underpaid cases in the re-review sample.

For example, the computation of this relationship for ineligible cases (called the regression coefficient "b") is as follows:

$$b = \frac{p_a - (p_f)(p_{n'})}{p_{n'}\ (1 - p_{n'})}$$

where:

p_a = the proportion of cases in the Federal subsample in which there is Federal/State agreement on ineligibility

p_f = the proportion of cases in the Federal subsample which are ineligible based on the final Federal determinations (after resolution of differences)

$p_{n'}$ = the proportion of cases in the Federal subsample in which the original State finding was ineligibility (irrespective of Federal findings)

The above value (b), once calculated is multiplied by the difference between the State original findings in the Federal subsample ($p_{n'}$) and the State original findings in the State full sample (p_n) (the former subtracted from the latter). The product is then added to the final Federal finding in the Federal subsample (p_f) to produce the regression estimate of case rate of ineligibility ($\hat{p}_f$). The formula is as follows:

$$\hat{p}_f = p_f + b(p_n - p_{n'})$$

The regression case error rate has the following properties:

- When original State findings agree with final Federal findings in all cases in the Federal subsample, the error rate computed from the State full sample becomes the "official" error rate.

- When there is no discernible relationship between final Federal and original State findings in the Federal subsample, the Federal case error rate computed from the Federal subsample becomes the "official" error rate.

- When the relationship is good but not perfect between the Federal and original State findings in the Federal subsample, the "official" error rate computed by the regression formula method would be close to the error rate computed from the State full sample.

These properties clearly illustrate the logic and reasonableness of the use of this method. The initial State findings are used in such a way in the Hansen formula that if Federal and State findings are in perfect agreement with each other in the subsample, then the regression error rate will equal the initial State error rate and the result is an indication of the validity in the State QC findings. If, on the other hand, the relationship between Federal and State findings is poor, the regression rate will equal the final error rate in the Federal subsample and the result shows little validity in the State QC findings.

The same principle applies in the use of the regression formula method for computing payment error rates except that, in addition to the comparison of Federal and State error payments, total payments in the Federal subsample are regressed against total payments in the total caseload to further refine the data. (Standard errors for each error rate involves the calculation of correlation coefficients.)

Operational Problems

As all of you know, no survey is without abnormalities and the AFDC-QC system is certainly no exception. Imagine administering a sample survey from 54 separate jurisdictions, each having its own eligibility and payment rules and regulations. On top of that, imagine administering a subsample from 10 Federal regional offices.

Nationwide, there are over 1200 State QC staff involved with the sampling, review process, and the data and program analysis of the quality control review in the States--each with varying amounts of skill, training and ability. Problems come up all the time which could affect the comparability of one sampling plan to the next, and ultimately the validity of the data from one State to the next. For example, take the case of stratified sampling. Theoretically, this should not cause a problem, as this is not a particularly sophisticated technique. We must ensure that all States with similar size caseloads have the same precision for a fixed error rate. However, in desiring estimates for geographic areas which contain relatively small proportions of the total caseload, some States allocate the stratified sample inefficiently--i.e., in such a manner as to increase, rather than decrease, the sampling error. In order to keep the precision within our standards, these States must increase their total sample size. This can significantly increase costs for the State.

Another complexity we have involves State payment dates and sampling frames. Everything runs smoothly when the State issues monthly payment checks on the first of each month and selects sample cases from one central payroll listing covering all cases. In fact, however, the quality control system must accommodate all State systems designed for paying welfare clients. For example, one State has no central payroll file at all but issues checks from 30 separate locations; another State issues semi-monthly checks, not always in equal amounts; a third State has 20 different payment dates in a month--semi-monthly payment dates arranged according to the alphabetical order of names on the assistance rolls; a fourth State issues one check a month to cases which include earned income and two checks a month, again with scattered payment dates depending on last name of clients, for all other cases.

All these differences in payment structure cause complications. The correction of a payment error in the review month of sample cases after they have been selected as part of the QC review could conceivably result in all sample cases being found to be correct. Therefore, States making payments on a semi-monthly basis are required to delay the selection of their sample until after the second semi-monthly payments have been authorized.

Another complication involves cases for which a review cannot be completed because the client has moved out of State, or is unwilling to give information, or cannot be located. Such cases essentially amount to nonresponse but are usually a very small part of the sample. A problem usually comes up when such a case appears in the Federal subsample.

Since the Federal subsample is an unbiased estimate of the State QC sample, both samples must consist of the same type of cases. If cases dropped by the State are completed by the Federal re-review, States can accept the Federal findings or go out and do their own review. If, on the other hand, a Federal reviewer cannot complete a case already reviewed by the State, the State finding on the case is assumed to be correct. While we realize that this procedure tends to inflate the correlation between Federal and State findings, dropping such cases from both Federal and State review would lead to an estimate that is not representative of the total caseload it seeks to estimate. This situation arises because of the difference in time of the State and Federal reviews. (The Federal re-review generally takes place anywhere from 30 to 60 days after the State has completed its review.)

One of our biggest problems, however, has been with States wanting to change their initial findings after they have been submitted to the Federal regional office, particularly at the end of the six-month reporting period. As indicated earlier, the regression estimate of the error rate makes use of the relationship between final Federal subsample findings and initial State findings in the full State sample. Since the initiative for making changes rests with the State, we have no assurance that all changes, whether or not they favor the State, are identified and reported. Although the Federal reviewer can validate changes initiated by the State, failure to report all changes can bias the final results since the universe for the Federal subsample is the State full sample. It is for this reason that regional offices are not permitted to accept State changes after results of completed reviews are submitted.

Conclusion

The AFDC-QC system, in conjunction with other administrative tools, has made a substantial contribution to the improved management of the AFDC program. Since 1973, case rates of ineligibility, overpayment and underpayment have been reduced 35 percent--from 41.1% to 26.7%--resulting in a cost avoidance savings of almost a billion dollars in State and Federal funds. The system makes use of the accepted objectives, principles, and techniques of statistical quality control. It is an ongoing management process using sample inspection, providing continuing data on error rates and identifying nature and types of error for guiding corrective actions for error reduction (whether the error results in overpayment or underpayment).

EVALUATION OF THE FOOD STAMP PROGRAM

Stephen J. Hiemstra, U. S. Department of Agriculture

Program evaluation involves the systematic study of costs and benefits received from a program within the constraints of achieving the objectives as specified for the program. Any such evaluation must begin by laying out the pertinent objectives and measuring the extent to which they are being achieved. That in itself is often no small task. Program objectives tend to be conflicting and subject to evolution over time as legislation is amended and interpreted.

This paper outlines my concept of the primary objectives of the Food Stamp Program, indicates the extent to which the program has been evaluated in light of those objectives, and discusses the specifications for a new major evaluation of the program that is needed.

I. Program Objectives

The objectives of the Food Stamp Program are legislatively two-fold: (1) to increase the demand for farm products, and (2) to increase the nutritional levels of low income people. These two objectives were intended to be largely complementary, or at least compatible. However, that isn't necessarily the case. The program in its construction allows freedom of choice to recipients to purchase the kinds of foods that meet their personal preferences. Increased expenditures for food don't necessarily mean that the foods are any more nutritious.

There has been considerable shift in emphasis given to these two objectives since the original Act was passed in 1964. Initially, a great deal of emphasis was placed on expanding the demand for farm products. But there has been considerable shift in emphasis since then toward increasing nutritional levels of low income people. To some extent the program itself has been redesigned to give this second objective more emphasis. Also, there has been little need to stimulate artificially the demand for food in recent years of rising food prices and large foreign demands.

A third implicit objective of the program has also been getting increased attention in recent years and to a considerable extent overshadows the other two objectives in the recent debate on Food Stamp Program reform. I am speaking of the transfer of resources that takes place as a result of participation in the program. There is much concern about the equity considerations surrounding the redistribution of income brought about by the program. This is expressed in the policy debate concerned with (1) the appropriate level of eligibility for the program, (2) the amount and size of the contributions, if any, required of participants in the Program (purchase requirement), (3) the appropriate definition of income and income exclusions to allow in determining eligibility and benefit levels, (4) the levels of participation in the program in relation to the number eligible, and the outreach efforts that have been conducted.

Rather surprisingly, these concerns for equity and redistribution of resources have tended to overshadow concerns about meeting the food and nutrition objectives of the program. One of the emerging policy issues, however, highlights the conflict in this implicit objective of income redistribution with the food and nutritional objectives. That issue is whether or not the program should continue to have a purchase requirement or whether the benefits should simply be given to the recipient at no cash outlay to him. At issue is the importance of providing the opportunity for the recipient to purchase a nutritionally adequate diet vis-a-vis the importance of increasing participation.

The purchase requirement is designed to insure that the program has impacts at the margin of previous levels of food spending and thereby raise food consumption and nutritional levels of recipients, and in aggregate expand the demand for farm products.

II. Existing Data and Past Research Efforts

Over the years a number of surveys and studies of effectiveness of the Food Stamp Program have been conducted. These include the following:

1. There have been two Nationwide surveys of the profile of participants in the program. The first one, with data as of November, 1973, was conducted under contract as a household survey. The second survey was an in-house activity by our own field staff based on information in case files of households certified for the program as of September, 1975. It was gratifying to see that the results of these surveys were compatible even though they utilized different methodologies.

2. There have been several pilot studies of food and nutrient consumption by program participants in a given county or locality, but no such studies have been National in scope. Most of them involved obtaining a sample of households participating in the Food Distribution Program just before it was terminated, and a similar number of households eligible for that Program but not participating. Then, about six months after the area had transferred to the Food Stamp Program surveyors went back and reinterviewed as many of the original households as could be found. Four-way comparisons of food consumption patterns of the households were analyzed.

Such studies were conducted in the early years of the program in Detroit, Michigan, and Fayette County, Pennsylvania; Detroit, Michigan; and Sunflower County, Mississippi. More recently, studies have been made in two counties of Southern Pennsylvania; Kern County, California; and Bullock County, Alabama. The earlier studies utilized 7-day recall of food use, patterned after the Nationwide Food Consumption Surveys of the Agricultural Research Service conducted most recently in 1965-66. The recent studies in Pennsylvania, Alabama and California, however, involved 24-hour recall of food use.

These studies all contained a large number of data and methodological problems. With only two points in time spanned, cross-sectional comparisons of a relatively small number of households, and a myriad of foods used, there obviously were more variables affecting the data than could be explained statistically by the degrees of freedom present. Variability in consumption of many food items due to the uncontrollable variables appeared in many cases to outweigh the differences in consumption attributable to the introduction of the Food Stamp Program. At best, only tentative conclusions could be drawn from these pilot studies.

3. An overall assessment of the impact of the Food Stamp Program in terms of food expenditures at the retail level was conducted by the Economic Research Service of the Department. It was a synthesis of many of the previous studies that had been conducted. Its basic finding was that somewhere between 50 and 65 percent of the bonus value of the Food Stamp Program likely is reflected in increased food expenditures at the retail level (MPC of .5 to .65).

In going a step further, it pointed out that about two-fifths of that amount likely was reflected in increased value of food as purchased at the farm level. That conclusion, however, was based on average rather than marginal propensities to consume, and based on secondary data. It did not reflect marginal analysis of the Food Stamp Program itself.

4. A fairly comprehensive study of the Food Stamp Program was submitted to the Congress by the Department as a response to Senate Resolution 58, in June 1975. To a large extent this study compiled all of the evaluation studies that had been conducted up to that time. It pulled together data from many secondary sources as well as came up with new estimates on the number of people eligible for the program in relation to the number that were participating. A Supplement to that report also contained, for the first time, five-year projections of program size and cost. In addition to evaluating the program, the report included the first statement of the Department's recommendations for Food Stamp Program reform.

III. Need for Further Research

The very size of the Food Stamp Program (nearly $6 billion reaching nearly 18 million people) and its recent growth in the past two or three years has raised serious questions as to its effectiveness in meeting objectives. The need is clearly for National data that can be used to answer important questions of policy.

The question about the usefulness of the purchase requirement is a case in point. Cost effectiveness of this in-kind program as opposed to a cash supplement program such as a negative income tax or income security type of program is another issue. Related to these issues are questions about the duplication of benefits of the Food Stamp Program vis-a-vis other in-kind food programs such as School Lunch and the WIC (Women, Infants and Children) Program as well as various other Federal programs.

Analytical objectives designed to provide information relative to these policy questions include the following:

1. Determination of changes in food expenditures and food consumption of recipients that are attributable to the program, and that can be contrasted with studies of cash supplements.

2. Measurement of the extent to which recipients in fact receive a nutritionally adequate diet.

3. Assessment of the impacts of additional food demand generated by the Program upon aggregate food markets.

4. Examination of the importance of food stamps in relation to total food availability to the household, and in relation to other Federal transfer programs.

IV. Specifications of Needed Survey Data

With these objectives in mind, the Department for the past two years has been working on the study design of a major survey that would collect primary data from a National sample of households with the following specifications:

1. Data must be longitudinal in order to minimize variability due to factors other than Program participation, and to measure the impacts of dynamic variables.

a. Cross-sectional analyses that have been conducted in the past have not been successful in isolating the impacts of the many variables affecting food consumption of individual households. Typically, such household studies "explain" only relatively small parts of the total variability in consumption of individual food products.

Compounding this problem is the observation from several studies of factors affecting Program participation that nonparticipating "control" households often appear to have unique reasons for not participating in the Program. The recent Kern County, California, and Bullock County, Alabama, studies each listed about 10 reasons for not participating in the Program. The recent ERS study of participation came up with a different set of factors.

In sum, studies such as conducted in Kern County and Madden's study in two counties in Pennsylvania were not able to sort out statistically all of the factors affecting participation using cross-sectional data. Even though variations in consumption over time are also subject to a large number of exogenous influences, our hypothesis is that more variables would be endogenous to a longitudinal model than to a cross-sectional model. Actually, the intent is to analyze the data utilizing both dimensions of the data.

b. Longitudinal analysis can assess the impact of program participation, because of the considerable amount of variability in participation by the same households over a period of time. Consumer Population Survey data and an in-house survey have shown that over a 12 month period, there are somewhere between 40 and 70 percent more unduplicated households participating than participated in any single month of that same 12 months.

c. Similarly, longitudinal data can readily span a number of changes in program design and benefit levels over time because historically such changes have been rather numerous over a one or two year period of time. Price level adjustments are mandated twice a year in both stamp issuance and eligibility levels.

Price level adjustments are quite costly to the Government because the entire cost becomes an added Federal expense if household incomes remain unchanged. Nevertheless, no direct analysis of the impact of these incremental Program changes on food expenditures or food prices has been possible in the past. Such analyses as have been made have relied upon assumptions and secondary data.

2. The length of survey period of data collection for individual households must be long enough to overcome the expected normal variability in household expenditure and consumption levels. Since the assumption is to be made that all food brought into the household is used except for normal waste and loss, the period of observation must be long enough to even out the inventory problem. Major inventory changes, however, would be allowed for.

Data from the Atlanta Panel survey (1956-1962), which collected continuous data from the same households for six years, and special tabulations of data from the Market Research Corporation of America (MRCA) panel 1973 data provided the basis for specifying that the survey period should be 3 to 4 weeks long. Coefficients of variation were computed for varying time periods up to six week averages based on expenditures for selected food products. Variability decreased greatly as the length of observation increased up to four weeks but relatively little reduction occurred after that time.

The time period covered by sample data should match up with program operational characteristics, or the data must be adjusted to it. Food stamps typically are distributed on a monthly or semimonthly basis. Food purchases would be submitted on a weekly basis, which is thought to coincide with typical shopping patterns. The analytical adjustment to a monthly basis is complicated because of the varying number of days per month. For that reason, the calendar month perhaps is the best unit of observation.

3. Detailed quantity of food consumption and food expenditure data are needed to determine the nutritional adequacy of the diet of recipient households and to measure impacts upon markets for individual food commodities. Since nutrients vary considerably among foods within the same food group, and even among forms of the same food, careful specification of the data is a requirement for analysis.

a. Standard procedures developed by the Agricultural Research Service in analyzing their decennial household food consumption survey data will be used in converting quantities of food on an "as purchased" basis into nutrients, after allowing for loss of food in storage, conversion to edible basis in preparation, and plate waste.

b. Reliance upon analysis of nutrients in the food used is made in the context of the legislation that specifies the Food Stamp Program to be a food program which is only one--albeit an important one--aspect of a total health or medical program designed to impact upon the total nutritional status of recipients.

The Food Stamp Program is designed to operate through the free marketing system, allowing free choice in the selection of foods. Since objectives are to increase food use and nutrient levels of low-income people, there are implied assumptions that, first, there are food and nutritional problems to solve, and, secondly, that they will be solved by providing increased purchasing power to recipient households.

c. The household is specified as the unit of observation. This decision also reflects program design that provides increased purchasing power to the household and leaves it to that household to maximize internal food use. Nutritional status of individuals within the household would not be assessed. Analysis by individual household sizes--with particular attention given to large and small sizes that weigh heavily among Program participants--is contemplated.

d. If there is in fact no measurable nutrient impacts of the program, it may be due either to the lack of a substantive nutritional problem to be solved, or a poor delivery system for solving it. For example, recipients might be spending additional money on nonnutritious foods, as is often charged by program critics. It is beyond the scope of this study to assess the total nutritional status of the population. That is the job of the HANES Survey and related efforts.

e. Detailed quantity and expenditure data are also needed to relate food purchases to the size of the markets for individual food commodities to assess the impacts of the Program upon U.S. market demand.

4. Diary tabulation by cooperating households is to be specified as the form and method of the data collected. There is limited statistical basis to choose between recall and diary data. A recent methodology study by the Agricultural Research Service of five or six different lengths of recall and diary periods showed that recall data generally tended to show higher rates of food consumption than did diaries. No assessment was possible of the true mean, however. If diary data underreport the mean, there would be a bias introduced in assessing the extent of the program achieving a nutritionally adequate diet. Therefore, every effort will be made to stimulate complete and accurate reporting.

V. Sample Design

The sampled population should consist of all households participating, likely to participate, and eligible to participate over the period of observation in the Nation. Unfortunately, no such listing of households is available nor can one be readily compiled. There are listings of program participants available at the local project level but it would be difficult to add files of potential participants.

For that reason, an area probability sample must be drawn for screening purposes and enough information obtained to determine eligibility status. That will be cumbersome because eligibility is complicated to determine and income normally is subject to verification by the caseworker. Eligibility is complicated by a long list of income disregards that are available.

Nevertheless, there is substantive interest in the file of eligible nonparticipants. There is considerable interest in determining the number, profile and location of this group for purposes of outreach and assessment of the extent to which the program is reaching the intended group.

A complicating factor is the variation that is known to exist in monthly incomes of those potentially eligible to participate and the high rate of movement on and off of the program. It is not known to what extent such movement is due to change in eligibility status versus other reasons.

Sample size has not been determined, but it likely will be in the range of 2,500 to 3,000 households located in 60 to 65 PSU's defined to be SMSA's. Differential sampling rates within SMSA's may be allowed, and area clusters identified for efficiency in data collection.

Decisions are yet to be made as to the desirability of extending the study to Puerto Rico, Virgin Islands and the other Territories where the Program also operates. Conceptually, it should be extended but the sample would need to be stratified to cover expected additional variability in those areas.

VI. Data Collection Problems

There potentially are serious problems of respondent burden and consequent nonresponse from this type of data collection. For that reason, consideration is being given to limit the length of data collection from each household to the minimum necessary to meet the requirements imposed by expected data variability, analytical needs, and program capatibility.

Dropouts are expected to be a problem due to many factors--family splits, geographic movements, lack of interest, and reaction to the sizable burden imposed by the collection activity itself. Plans are to follow as many households or splits as practical and to subsample nonrespondents. But there would be no replacement of households for those withdrawing from participation. Respondents likely will be compensated for providing data to minimize the dropout problem.

VII. Analytical Problems

The survey will be designed and data collected to minimize analytical problems, but many of them will remain. These include the following:

1. Nutrient content of foods is quite variable. The study will assume average nutrients in specified foods; it will not employ chemical analysis because it would be impracticable in this size and type of data collection.

2. Nutrient needs of people are variable and dietary standards themselves are set at several standard deviations above average needs (which are only imprecisely known). Thus, failure to meet 100 percent of the nutrient standards may not be indicative of failure of meeting nutrient needs of individual households.

3. Length of time period for analysis will be a compromise between the desire for homogeneous data, availability of data, and variability in program participation status. Quarterly data may be desirable for minimizing problems due to lack of inventory control, but participation status changes on a monthly basis. The intent will be to measure the impact of multi-Federal program participation, but such status likely will also change on a monthly basis. Lack of matching time periods is complicated by sample data collected on a weekly basis.

4. Assessment of net impacts of the program will be complicated by the operation of many random influences affecting expenditure and consumption patterns that may not be quantifiable. Individual household data are known to be almost infinitely variable and only the major factors can be quantified. Statistical analysis utilizing the pooling of longitudinal and cross-sectional data is envisioned, since the data represent a time-series of cross-sectional sample points.

5. The time period of overall assessment will be limited to two years of observations--perhaps 6 to 8 sample points of quarterly data. Although considerable variation will be expected in the substantive variables, there may be some dynamic features of the program that will not change sufficiently to allow statistical analysis. For example, some kinds of people on the program will not change program status. Changes in program status are less frequent for public assistance and SSI than for nonpublic assistance households.

6. In addition, the analysis may be complicated by the aggregate income and employment, and food supply and demand picture confronting the program during the period of observation. Inflationary problems are to some extent endogenous to the system being studied, but it is expected that the other factors will be held constant through normal statistical procedures.

VIII. Plans for Related Studies

Aside from the household consumer panel, FNS plans to conduct the following additional studies in the near future:

1. Survey of household assets

This survey, also to be conducted on a National probability basis, is intended to obtain data needed to provide the Administration and the Congress with information regarding the asset holdings of food stamp households and of other low income households eligible for participation. Included would be questions covering both the value of assets currently excluded from consideration in determining eligibility and assets currently included in such determination.

If the household consumer panel discussed above is not approved, this survey would be expanded to provide the basic profile data expected to be generated by the screening survey regarding eligible nonparticipating households.

2. Survey of Level of Understanding of the Program

This survey will focus on determining the level of knowledge of the Program, for purposes of gaining insight into the nature of the reasons for nonparticipation of those households currently eligible for the program. It should provide understanding of the need for outreach and the direction that outreach programs should take in being effective. It will attempt to be more in-depth in scope than previous studies of reasons for nonparticipation have been in the past.

3. Study of Impact of the FSP on Indians

This study will try to provide insight into the lack of acceptance of the Program by many Indian Reservations. It will study areas where the Program has been implemented and focus on rates of participation, food prices charged, availability of food stores, costs and convenience of program operations, and profile of participants. To the extent possible such data will be matched against previous studies conducted while the Food Distribution Program was in operation, such as the study conducted on the New Mexico portion of the Navaho Reservation.

XI. Summary

Comprehensive evaluation of the Food Stamp Program has not been conducted to the extent that is needed for policy purposes. Several elements of the program have been studied carefully--particularly those issues related to the equity questions of program eligibility, participation, and size of benefits. But cost-benefit analysis has been lacking because of little comprehensive information available on the program impacts upon either recipients or aggregate food markets.

A longitudinal study of the food and nutritional impacts of the program is now being designed to provide the data for such an analysis. There are many statistical questions under study. They have been addressed by a study group which has recently given the Agency a report on this subject. Nevertheless, the methodological problems have not all been resolved at this time.

DISCUSSION

Robert Ferber, University of Illinois

All three papers presented at this session are very informative from the point of view of describing ongoing programs or plans for such programs. At the same time, they are somewhat frustrating to a statistician such as myself because hardly any statistical data are provided nor is there any analysis of the operations and effects of these programs. For this reason, and also because I have had no personal experience with these programs, my comments are more in the nature of questions for future consideration, and some may be naive.

The paper by Deutch is a good example of the quandry in which I am put. The procedures described in the paper seem excellent and a great deal of information is provided on how the SSA Measurement System is evaluated. However, we are given virtually no information on how it works in practice, perhaps because the system is relatively new. Thus, what are the results of this evaluation? What sorts of errors have been found? How have these results been used? To what extent have they resulted in any improvement in the SSA System?

Information of this type would be very useful, as would information on how well the system is operating. For example, how long does it take until the six-month sample is 95% complete and administrators are given the results.

A broader sort of question relates to Deutch's very apt opening comments that in quality control the production worker may be only part of the problem and that most of the problems may lie in the area of management and the structure of a system. This sort of observation might well be leveled at the SSA evaluation itself. In other words, in this system, to what extent are problems due to the production workers (in this case, the evaluators), and to what extent are they due to the evaluation system and its administrators?

The Ossman paper on the AFDS Program is also very informative with regard to the manner in which the program operates. Unfortunately, we are again given little information about what actually happens, which leads me to raise a number of questions, such as the following:

1. What is the rationale for using systematic random sampling in the re-review at the federal level of the cases reviewed by the states? Would it not be better perhaps to stratify in the re-review sample by amount of error reported in the state review? Would this not yield a better base for evaluation of the efficacy of the state reviews?

2. Who are the reviewers that check the cases at both the state and the federal levels? How thoroughly are they trained.

3. How much time is available for review of the case record? Is this time adequate?

4. In this review is it a good idea to put the main emphasis once more on the client as the principal source for information? Is the reviewer given, for example, information that the client had supplied previously (and which might therefore bias the reviewer's judgment), or does the reviewer work completely without any such information as seems to be the case with the SSA evaluation?

5. To what extent do more intensive reviews yield more accurate results? In other words, what is the trade-off between more thorough reviews and more accurate pinpointing of errors, against higher costs and more time?

6. How are the findings implemented in the form of procedural changes? Does the initiative come from the federal level or the state level? What sort of approvals are needed for such implementation to take place?

Numerous operational problems can be expected in a program of this type. It would have been interesting to hear more about these problems, their effects and how they are handled. For example, similar to interviewer effects in personal interview surveys, to what extent is there a reviewer effect in the review of these cases? What sort of scheduling and personnel problems arise in this work? What is the frequency with which different types of errors have been pinpointed? Only a couple of percentages cited, just enough to whet one's appetite.

The Hiemstra paper differs conceptually from the other two because they report on evaluation programs already underway whereas here we have an evaluation program still in its formative stage. Hence, discussion and interchange of ideas should be especially fruitful in this case. Moreover, to judge from the presentation, the planning for this study is by no means completed.

As Hiemstra mentions, the contemplated national survey would be of major proportions. The four objectives mentioned are certainly very broad and also very worthy. I would suggest a fifth, namely, some investigation of the attitudes of the people in the program and their experiences with it.

In view of the huge amount of data that will be sought, a number of operational problems will have to be handled successfully, such as setting up control groups, obtaining cooperation, inducing people to remain in these panels, and using various devices to stimulate complete and accurate reporting, among others. Since I am not familiar with some of these other studies of the Food Stamp Program, it would seem most useful in this brief space to offer some general comments on the procedure to be followed in carrying out this study.

In particular, I would urge that a study of this magnitude not be launched without a great deal of consideration to previous studies that have experimented with these techniques, and with some pilot work to test various alternative approaches. The 1972-73 national survey of consumer expenditures and income conducted by the U.S. Bureau of the Census for the BLS contains a gold mine of such information for these purposes. An evaluation of those data is currently being carried out by Bob Pearl and should be available in the next few months.

As a further preliminary step, pilot tests of particular techniques and combinations of techniques in limited geographic areas should be carried out in advance of the projected national survey. Experience with other major studies has shown, without exception that such pilot operations are invaluable for reducing costs and improving the efficiency of the later full-scale study.

In view of the uncertainty that seems to surround the planning of this program, it would seem like a very good idea for this agency to use a technical panel to assist it in all stages of this study, from the initial plans, through the implementation of the results. Even if an agency has highly trained technical survey people of its own, an outside panel can provide a different, and usually broader, perspective on the problem, and help lend credibility to the results.

From a more technical point of view, let me suggest that in a comprehensive survey of this type the data collection be divided into components so that the questionnaire is not too long. In other words, there might be a single basic questionnaire administered to all sample members, supplemented by several component parts, each part administered to a different interpenetrating sample.

Also, if the study can be designed as a panel study so that data are available on these households over time, the problem mentioned by Hiemstra of having many endogenous variables would be reduced very greatly in importance because most such variables that are endogenous on one basis are exogenous on the other basis. Hence, by using appropriate forms of multivariate analysis, problems of estimation because of the presence of endogenous variables would be of little importance.

SOCIAL SCIENCE RESEARCH IN THE POLICY ARENA

William A. Morrill, Department of Health, Education, and Welfare

My experiences in public life--particularly over the last six years--have significantly shaped my thinking on today's topic. These include a year as a deputy county executive in a large suburban jurisdiction in the Washington metropolitan area, a last tour at the Office of Management and Budget, and over three years at HEW as assistant secretary for planning and evaluation. I want to share some thoughts with you out of the accumulation of these experiences. These last three jobs that I have held seem to me to be at a juncture where I was deeply involved in trying to take what we know and what we believe and turning all that into what we should do--and in some cases, perhaps, what we should not do. And that hopefully, has given me some perspectives that may be helpful to you in your deliberations as one part of the group in charge of "what we know."

In getting ready for this discussion I thought about who's in charge of what we know and who's in charge of what we believe. We might say, rather simply that the social science researchers are in charge of what we know and that the politicians are in charge of what we believe. I may make more from time to time of that dichotomy, but we all know that it's really not that simple. I've yet to see anyone in the social research field who not only knew something, but didn't tend to believe something about what ought to happen next. And by and large those who are in charge of what we believe, at least hopefully, know something. But it is the interaction between these groups that has been an important part of my role. Somebody asked me at the table during lunch. "Does that make you a policymaker?" I'm not terribly sure that the answer is "Yes," but at least close enough to talk about it for a while.

In thinking about those interactions, it seems to me we are constantly faced with a series of paradoxes and I want to describe these a bit. In the course of this you'll find me returning to some familiar themes. For one thing, I am, unabashedly, strongly in favor of a utilitarian focus on the expansion of what we know, be it in the field of statistics and data acquisition, in social research, or the like. This is not in my view any attack on basic research or on the value of such endeavors, nor is it necessarily a kind of narrow philosophy that says, "I am only interested in today's problems." After all, a utilitarian focus is, in my view, also a process of playing that wonderful guessing game of trying to identify what issues are likely to be important enough five years from now to generate a consuming public debate. The problem is to figure out what I can do as a social science researcher or planner of such activities to get ourselves ready to conduct that debate in an intelligent fashion.

It also seems to me that while we have a responsibility for advancing what we know, there is an equal responsibility in these efforts--on both sides of any question--for some modesty about the likelihood of what we know at any given time being the crucial factor in decisions about what will actually happen. And whether you are in the role of policymaker or politician, on the one hand, or that of a social science researcher, on the other, this responsibility is still there. And we need to remind ourselves from time to time about these limitations.

Let me turn for a moment to what in my view those limitations are. There is first of all a problem of relevance. Now that's a subject about which we all talk a good deal--and not alway usefully. In some cases the limitations on social science research in the realm of relevance lie in what may be called the "interesting but not needed"category. Let me give you some illustrations of this out of my own HEW experience. Take for example the issue of sex discrimination. There has been some exceedingly important and needed research that has raised the general level of understanding about the nature of sex discrimination and how it affects our society. We welcome good research on tough problems, such as just what one should do about sex stereotyping in curriculum materials and textbooks in the educational enterprises of this country. But, we do not need beyond that an endless series of research efforts on such problems as, for example, the sex discriminatory provisions of the Social Security Act. They are there in plain view, and if they don't draw attention to themselves, or policymakers don't, the courts are quick enough to get at them. And we don't need to really explore very long to find out what we really need to do--as quickly as possible--to get rid of them. That's not a researcher's problem, but indeed a problem for those who are in the political arena. And I think we do need to be careful at all times not to engage ourselves in such "interesting but not needed" activity.

There's another equally and perhaps more difficult category of "interesting, but not very useful." It is my impression, for example, in the field of education statistics, that we have produced an enormous amount of data but, in many cases, not very much information on problems about school finance, what happening under the efforts of the country to desegregate, etc., and often an avoidance of what are clearly some of the predominant problems in the society around us. In this same category, of course, are the conventional problems of studies that are poorly done. But I would also call attention to the kind of social science research which concentrates excessively on factors which cannot--by activities of the society--be much manipulated. Let me illustrate that point. We know for example in realm of welfare that one of the causes, if you will, of poverty is family breakup. One result of longitudinal data studies is quite clear; namely, that a change in family composition can often drop families into poverty. What is important to find is that piece of information that is relevant in terms of designing what we do in the welfare field. But it is probably not worth a great deal of effort for policy to, say, get into the whole field of marriage per se and what's happening there since it is rather unlikely that our government is going to succeed, at least at the Federal level, in intervening in that kind of a problem. It is clearly, in our society, a problem that is not subject to governmental action. And yet there are dimensions of how a particular governmental welfare policy may impact positively or negatively on family breakup that could provide important pieces of information for action--and that distinction does need to be made. And this shows, in turn, the importance of asking the right question in the context of the governmental programmatic structure around us.

Another issue in the limitations of social science--and this one really has been for me a bit of a paradox--is found in the problem of popular values and biases, and what to do about these. One technique, of course, is to take that problem and, if one has evidence that the existing biases or values are at least open to question, to try and overwhelm them with the evidence. And indeed I have had some personal responsibility for that around the classic problem in welfare of what should we do about our work ethic and not to build programs that in fact would encourage people to withdraw from the labor market. But there probably comes a point where we need to consider the art of accommodation and to focus some of our research, not on trying to tell people that they don't believe the right things, but rather on how to accommodate both what the evidence shows us and those deep underlying value structures.

The only example that is perhaps most immediately relevant to some of the discussions that you are having in the health field is the issue of cost sharing in health insurance. Here we're wrestling with some enormous public biases and with the whole tradition in the insurance industry over some twenty or thirty years of first-dollar coverage, which the evidence suggests may not be the best way to handle the problems of national health insurance policy. We'll see over the months ahead just how that issue is going to play itself out.

There's also the problem in the social science fields of "no alternatives." We are rather good at looking at activities of one kind or another and demonstrating that they don't work. The area of manpower and education comes quickly to mind, in which a number of competent studies have suggested the disutility of many of those activities. And yet we are in an area of deeply felt societal perceptions and just saying that programs won't work--without discovering why they don't and giving better interpretations of our findings that can potentially suggest what our analyses show us about alternative ways to proceed in the area--does limit the value of social science research in the policy arena.

Lastly, there is the problem of the nonquantifiables. The literature is replete with the difficulties of capturing cost and benefits. But we need to remind ourselves very often that the problem as perceived is broader than just what we are able to quantify. Issues such as human dignity, health status, and quality of care are for the most part nonquantifiable. We must be cautious that those nonquantifiables are taken into account early because they surely will be when the subject gets to the decisionmaking level in the political arena. Illustrations of this could go on endlessly, but let me just deal with two of them. I think one of the problems of income maintenance reform in this country has been our inability to articulate properly the relationship between that activity and what might be called social services. And the inadequacy of much of our research data in that field has led us in turn to practice of trying to make improvements while ignoring an important dimension. This either prevents a presumably better notion from getting implemented or causes other problems. Some of the issues of relating social services to Supplemental Security Income program are but a recent illustration thereof.

Also in the area of standards, we are at the moment in Washington grappling with standards of all kinds imposed from the federal level which in fact are just proxies to try to achieve something we call "quality of care" for day care, long-term care, and the like. As a result of our inability to deal properly with that quality-of-care issue, we have instead acquired a whole set of proxy measures--and some undesirable and unintended consequences. We now have penalties on providers for inadequately following the proxy measures of quality--and in turn we have penalized the recipients of the service.

In all of these issues I think there is a responsibility for everyone in the field to be mindful of. All these limitations can't be observed at all times; I think the opportunities and achievements now and prospectively are substantial. I think we are becoming increasingly sophisticated, both among those who are expanding the state of our knowledge and among those who are trying to do something with it. Some of these activities over the last four or five years have really made a difference in what has happened. If one looks at the role of experimentation, both the income maintenance and the health insurance experiment--prospectively in the latter case--I think have changed clearly the nature of the debate, not making it easier but perhaps focusing it better. What we have learned about labor response, what we have learned about accounting periods in the income maintenance area, what we can prospectively learn from the health insurance experiments about the relationship between demand and price and hopefully about the relationships between health insurance coverage and health status are important possibilties that can and will affect public debate.

In the role of surveys and data acquisition we've seen a remarkable event in the enactment of the recent education amendments where there was a heavy debate within the House and Senate committees over an endless series of data runs on how the allocation of education dollars would affect their perceived objectives. And they used that information extensively in coming to decisions about the proper allocation of resources. Perhaps the final choices were not those that an idealist might follow, but the fact that data had an impact is, I think, unquestionable. And indeed other legislation has led increasingly to a surprising and detailed imposition by the Congress by statute of new demands for information. The survey of income and education to provide more detailed and needed data on children in poverty is but one illustration. We are conducting a survey of institutionalized persons to get at various issues of disability. There is a health financing survey which is a demonstration of some remarkable cooperation internally within HEW among its constituent (and occassionally warring) parts to do something that was really important in gaining some new data. The income survey development program for which my office is providing some leadership is, prospectively, another example.

And one moves from surveys and data into straight-forward policy analysis modeling activities and the like where we are getting increasingly better and our product more wanted in examining at least the consequences of possible courses of action with more clarity. And yet in all of this there is a sense of dissatisfaction in many ways on both sides, both by the performers and the potential recipients. As I recently learned with some pain in the recent action of the Senate appropriations committee on my own policy research activities, there is a point of view of saying we are not terribly sure what this is all good for. "You folks out there spend a lot of money on these esoteric research projects and never tell us anything we really need to know."

I think clearly there is a call for a more structured dialogue between those who are trying to expand our knowledge base and whose are trying to do something with it. Between policymakers, policy analysts, researchers, data gatherers, and the like. And that dialogue, it seems to me, has some constituent requirements that I'd like to suggest. While meetings of this kind within professions are important devices for exchanging what's happening and what might go on in the future and learning from each other, there is an equally strong demand I believe for cross-professional contacts. And more listening for cues from other disciplines. And I confess I am as guilty as others in tending to get into a kind of narrow position of listening to the professionals with whom one associates as a normal matter and not listening to that less distinct voice from off-stage which can lead one to recognize some very important things about what you are doing or ought to be focusing attention on. We are trying in HEW, perhaps more than before, to provide for more structured interaction in our inter-agency activities. And many of the above illustrations I have used are important. This is in fact an effort to reduce a bit of the paranoia about each other between those who were using information and those who are providing it. As the chairman suggested earlier, perhaps some initiatives on both sides to reach out across those barriers and say "look its just not the other fellow's responsibility, we need to do something about that ourselves."

And for those of us on the policymaking side it involves perhaps most critically a commitment of time which is perhaps the most precious commodity of top decisionmaking, to sit down periodically and listen to what's going on in the field and ask ourselves perhaps a series of questions about our research and other relatd social activities and about the questions we are asking. If we work on a particular problem and if we find the answer will it make any difference to what might happen? If I find the answer will that answer be more likely to raise more questions than I have provided responses to--thus raising an issue of whether I've asked the right question in the first place? If the questions are right, can I find the answers in some timely way or before the political process moves on to make a decision-with or without our findings? And am I working mainly today's and yesterday's problems, or am I thinking sufficiently enough about those problems with which our society is going to grapple five years from now?

And with that set of tough questions, let me stop orating here and let me take on your questions if you have some for the remaining time.

DO OUR STATISTICS MEASURE THE REAL LABOR MARKET HARDSHIPS?

Sar A. Levitan, Center for Social Policy Studies, The George Washington University
Robert Taggart, National Council on Employment Policy

The Anachronisms of Labor Market Statistics

It is critically important that the economic indicators we compile and use in decisionmaking accurately portray reality. They should be valid and reliable measures of the factors that genuinely reflect the state of society. Current labor market measures, developed at the end of the Great Depression, reflect the primary concern of the time--the availability of jobs for those able and willing to work. Since then the unemployment data have proven to be a valuable multi-purpose indicator. The data were applied as a measure of the available stock of unused human resources in the economy. Based on the unemployment concepts the Phillips curve formulation suggested a relationship between labor market slack and inflation. Unemployment was applied as a basic variable in describing and predicting individual behavior. Before an elaborate transfer system was developed, the unemployment rate also served as a reasonable indicator of economic hardship.

Changing Realities

Over the decades, changes in the labor market and in our society have eroded the validity of the unemployment measure as an economic indicator for policy determination. Changes in the structure of the domestic economy have apparently altered the relationship between unemployment, wages, and inflation. Labor market developments were not primarily responsible for price changes in the 1970s. An ever increasing share of the labor force comes from households with two or more earners.

Idleness is increasingly an acceptable and voluntary option whose impact is softened by transfer payments or by multiple family earners. Many workers claim they want jobs but are only half-heartedly looking. More would take jobs if working conditions were favorable, but they are not actively seeking work. Others may feel it prudent and possible to look longer in order to find a higher paying job. Extended unemployment compensation, welfare, food stamps, social security, veterans benefits and other aid reduce earnings losses and even generate work disincentives; recipients with no interest in work may claim to be able, willing, and actively looking solely to meet program requirements. Finally, for workers trapped in a "secondary labor market," intermittent employment is the product of low wages, bad jobs, and employment situations in which turnover is accepted and even encouraged.

As a product of these changes, the relationship between unemployment and hardship has been increasingly obscured and unemployment statistics are no longer valid measures of economic and social health they once were. Joblessness among teenagers rarely affects the well-being of families. Many unemployed have a spouse with very adequate earnings, or else the family may have other income sources. Persons who do have intermittent work may have low earnings over the course of a year and even full-time work is no guarantee against poverty where there are many mouths to feed. Many full-time and intermittent workers end up with a lesser income than the families of the more affluent unemployed.

The Need for New Numbers

The shortcomings of the unemployment rate and other official labor market statistics have not gone unnoticed. Questions raised in the 1950s led to the appointment by President John F. Kennedy of a blue-ribbon panel, the Gordon Committee. The panel reviewed the concepts and suggested some revisions, but paid little attention to the emerging impact of income transfers upon work. No effort was made to redesign the unemployment data as a measure of economic hardship. Later, in the 1960s boom years, there was concern with persisting structural problems in ghettos, depressed areas, and among racial minority groups. The Great Society's policymakers, convinced that statistics then understated the seriousness of social problems, believed that a new measure of hardship would justify its active welfare efforts. When administrations and economic conditions subsequently changed, the statistics were questioned from another perspective. Some argued that the prevailing level of unemployment did not reflect economic deprivation as it had in the past, and that a higher rate had to be accepted to achieve price stability. The relevance of unemployment data as hardship measures was further questioned in the mid-1970s when unemployment compensation was liberalized and expanded, and when massive joblessness was greeted by public indifference.

Reacting to these concerns, the Bureau of Labor Statistics made efforts to expand the scope and coverage of the Current Population Survey data. Weekly earnings data were compiled each May and data on discouraged workers were added quarterly. An expanded household survey added details about family status and persons outside the labor force, but BLS made little effort to change the concepts that underlie the collection and presentation of unemployment data. The support for a hardship measure dissipated as unemployment became more widespread and policymakers showed little interest in a measurement which would accentuate labor market pathologies. The idea of a weighted unemployment rate to measure unutilized human resources or to justify raised unemployment targets became less pertinent when joblessness approached depression levels. Unemployment compensation benefits expanded as a stop-gap measure but critics claimed that many claimants were not seeking work. As in the past, debate over the numbers abated as conditions improved. The supremacy of the long-standing concepts was demonstrated when President Gerald Ford touted as part of his economic record a few downward ticks in the unemployment rate from the highest levels since the Great Depression, and his critics were sidetracked in a fruitless debate over the feasibility of an arbitrary full-employment goal.

The unemployment rate is like the proverbial shoe: we wear it because it is familiar even though it has become disfigured and the sole wears thin. We debate minor changes in joblessness and faraway targets without really knowing what the numbers mean. We continue to ignore the realities of a drastically expanded transfer system which provides some support to at least one of every four Americans. We still think in terms of neoclassic supply and demand theory despite the demonstrated interrelationships between low wages, discrimination, welfare and unemployment in the secondary labor market. In brief unemployment rates and other official labor market statistics have become inadequate to explain the ever changing labor market conditions. New concepts and new measures are needed for public policy formulation. Multibillion dollar programs and new proposals regarding employment policy may be riding on misconceptions about labor market operations which are based on Current Population Survey statistics.

A Hardship Measure

With earnings the predominant and societally-preferred source of income, a crucially important concern is the labor market's ability to provide workers not just a job, but a minimally adequate income. The long-term unemployed are likely to live in deprivation because of their earnings loss, but others besides these unemployed are failed by the labor market. Part-time employees seeking full-time work, intermittent workers, persons withdrawing from the labor force because of limited job opportunities, and, of course, low wage earners may all have deficient incomes. But many individuals with similar employment problems may not face economic hardship if there are other earners in their families or if they have alternative sources of income.

Concepts

The Employment and Earning Inadequacy index attempts to count all persons in the labor market who face employment and income problems. The prevalence of employment problems is first assessed by a "subemployment" measure defined to include the unemployed, discouraged workers not in the labor force who currently want a job but are not looking because they think no work is available, employed household heads who earned less than a poverty level wage in the last year (including those working full-time full-year as well as those working intermittently), and persons employed part-time involuntarily because of shortened work-weeks and other economic reasons. Full-time students age 16 to 21 years are excluded since they presumably are occupied in socially useful activity and therefore seek only part-time jobs, and since their income needs are frequently met from nonwork sources. Persons age 65 and over are also excluded since public pensions are now nearly universal and private pensions are much more widespread, reducing needs and labor force attachments. Only family heads are counted in the two low-earnings categories because other family members may frequently have only a peripheral attachment to the work force and hence limited earnings. [The technical flaws and conceptual difficulties involved in the proposed index were spelled out by the authors in Employment and Earnings Inadequacy: A New Social Indicator (Baltimore: The Johns Hopkins University Press, 1974), pp. 39-45, and in the Monthly Labor Review, Oct. 1973, pp. 24-27. Particularly troublesome is the distinction in the treatment of low earning males and spouses. The problems can be corrected when a more refined measure is developed.]

Despite the difficulties they face in the labor market, some of the subemployed may have an adequate personal or family income. In order to screen out these cases, an upper income adequacy test is applied. All persons whose family income for the preceding year was above the mean for families are excluded. The same holds for unrelated individuals with income above the mean. Since wide variations exist between metropolitan and nonmetropolitan areas separate mean incomes are applied to residents inside and outside metropolitan areas.

The Employment and Earnings Inadequacy index is calculated as a ratio of the subemployed with below-average incomes to the number of persons in the labor force, defined to include discouraged workers. The index indicates the proportion of people working, seeking work, or discouraged from seeking work who are unable to secure a minimum income and are also not fortunate enough to have other working family members or sources of income which ameliorate their own earnings problems.

Derivation

In March 1974 the civilian noninstitutional population numbered 148.2 million persons age 16 years and over. A total of 89.6 million were in the labor force and 585,000 were nonstudent, nonaged discouraged workers. The adjusted labor force was the sum of the two--90.2 million (Table 1).

Subemployment was the sum of five categories:

1. Unemployed. The Current Population Survey counted 3.9 million unemployed workers in March 1974 after subtracting students age 16 to 21 years old and individuals age 65 years and over.

2. Discouraged workers. There were 585,000 persons wanting a job currently but not looking because of discouragement over the prospects.

3. Fully employed low earners. There were 1.8 million family heads and 293,000 unrelated individuals who worked full-time, full-year in the previous 12 months and yet did not earn enough to reach the poverty threshold.

4. Intermittently employed low earners. Another 2.6 million employed family heads and 1.1 million unrelated individuals who had worked intermittently during the preceding year did not earn a poverty level income.

Table 1. Derivation of employment and earnings inadequacy index for March 1974 (thousand persons)

	Subemployed in Current Population Survey	Persons in households with above-average incomes	Employment and earnings inadequacy
Current Population Survey labor force	89,616	------	------
Discouraged workers (less students age 16-21 and persons age 65 and over)	+ 585	------	------
Adjusted labor force	90,201	------	90,201
EEI components			
(1) Unemployed	4,755	------	------
Less students age 16-21 and persons age 65 and over	- 866	------	------
Adjusted unemployed	3,889	-1,371	2,518
(2) Net discouraged workers	682	------	------
Less students age 16-21 and persons age 65 and over	- 97	------	------
Adjusted discouraged workers	585	- 133	452
(3) Employed full-time, full -year at less than poverty earnings (less students age 16-21 and persons age 65 and over)	2,076	- 179	1,897
(4) Employed intermittently at less than poverty earnings (less students age 16-21 and persons age 65 and over)	3,702	- 240	3,462
(5) Employed part-time involuntarily at less than poverty earnings	2,309	------	------
Less students age 16-21, persons age 65 and over, and persons counted in item 4	- 311	------	------
Adjusted employed part-time involuntarily	1,998	- 814	1,184
Total	12,250	-2,737	9,513
Subemployment and EEI index	13.6%	------	10.5%

Source: Tabulations based on Current Population Survey data.

5. Involuntary part-time workers. There were 2.0 million persons working part-time involuntarily for economic reasons who were not students, were less than 65, and were not counted among the intermittently employed low earners.

Adding these components, there were a total of 12.3 million subemployed in March 1974 out of the 90.2 million in the adjusted labor force, yielding a subemployment rate of 13.6 percent. Among these were 2.7 million persons living in households with above average incomes in the preceding year and therefore with questionable needs. Eliminating these from the subemployed left 9.5 million with inadequate employment and earnings. The EEI index was, thus, 10.5 percent.

Employment and Earnings Inadequacy--1974

Because of the very severe recession, the EEI figures for March 1975 are not representative of post World War II experience. Comparison of conventional unemployment data and EEI would be distorted by the deep 1975 economic slump. However, conditions in March 1974, when the unemployment rate was 5.3 percent, were more representative of post World War II experience. A study of the contrasts between CPS and EEI data should yield some insights about the potential value of the proposed measurement.

Components

Though unemployment substantially exceeded 1960 levels, the unemployed accounted for only a fourth of all persons with inadequate employment and earnings in March 1974 (Table 2). More than a third of the non-

Table 2. Components of the EEI index, March 1974 (thousand persons)

	Subemployed	Percent screened out	EEI	Percent of sub-employed	Percent of EEI
Total	12,250	22.3%	9,513	100.0%	100.0%
Unemployed	3,889	35.2	2,518	31.7	26.5
Discouraged workers	585	22.7	452	4.8	4.8
Low-paid fully employed heads	2,076	8.6	1,897	16.9	19.9
Intermittently employed heads with less than poverty earnings	3,702	6.5	3,462	30.2	36.4
Employed part-time involuntarily	1,998	40.7	1,184	16.3	12.4

Source: Tabulations based on Current Population Survey data.

student, nonelderly unemployed were members of households with above-average incomes and were not counted in the EEI index. Two-fifths of the involuntarily part-time workers were also screened out by the upper adequacy standards.

The low-paid fully employed heads accounted for a fifth of persons with inadequate employment and earnings, while the intermittently employed represented a third. The size of these low-earning groups is explained by several facts. Poverty among full-time working heads results from a combination of low wages and large families; intermittency compounds these difficulties by adding periods with no earnings. Many of the unemployed were affected by two or more spells of joblessness. Where the household heads earned less than poverty wages, it was very rare that earnings of other family members or alternative sources of income lifted the household to an above-average income. In terms of numbers, then, low earnings and intermittent employment accounted for twice the hardship as unemployment.

Employment Problems and Income

The EEI counts all persons with labor market problems and then screens out those who do not have severe income needs. This screening out process is vital in order to measure labor market related economic hardship.

The unemployment rate alone is not a very good measure of hardship. The unemployed in March 1974 (less students and the elderly) had a mean household income in the previous year of $11,443, or only 15 percent less than that of the total labor force (Table 3). The average

Table 3. Income and poverty status of the subemployed and persons with inadequate employment and earnings, 1973

	Subemployed	Persons screened out	Persons with inadequate employment and earnings	Incidence of poverty among subemployed	Incidence of poverty among persons with inadequate employment and earnings
Total	$ 8,446	$19,158	$ 5,364	31.9%	41.1%
Family heads	7,747	18,715	5,728	34.6	41.0
Wives	12,781	20,859	8,305	7.5	11.7
Other relatives	14,300	23,432	7,648	15.7	27.7
Unrelated individuals	2,926	9,415	1,933	59.2	68.3
Males	8,510	18,350	5,595	31.7	41.1
Females	8,441	20,204	5,157	32.1	41.0
Whites	8,958	19,411	5,477	28.4	37.9
Blacks	6,419	16,291	5,006	45.2	51.6
Metropolitan residents	8,833	20,360	5,552	31.2	40.0
Nonmetropolitan residents	7,777	17,149	5,030	33.1	42.8
Unemployed	11,443	19,844	6,869	16.9	26.1
Discouraged	10,057	22,010	6,540	26.4	34.1
Fully-employed low earning	5,898	17,534	4,800	50.8	55.6
Other low earning heads	4,851	17,026	4,007	51.5	55.0
Involuntarily part-time employed	12,566	23,094	7,750	6.7	11.3

Source: Tabulations based on Current Population Survey data.

income of the unemployed excluded in calculating the EEI was $19,844. This would hardly qualify in anyone's book as hardship. The discouraged and involuntary part-time workers also included many with dubious needs.

The screening process was especially important for wives and other relatives. Many of the unemployed were secondary jobseekers in families with substantial incomes. On the other hand, unrelated individuals with employment problems were more likely to face hardships because there was usually no one else to support them. Overall, the use of an income screen yielded an average annual EEI income of $5,364 compared with $8,446 for the subemployed. The proportion in poverty for the two groups was 41 and 32 percent, respectively; in contrast, only 17 percent of the unemployed were poor.

Who Bears the Burden?

The incidence of inadequacy varies significantly among different groups (Table 4). Some

Table 4. Characteristics of subemployed and inadequately employed, March 1974

	Percent of subemployed	Percent subemployed screened out	Percent of EEI	EEI rate
Family heads	51.6	15.5	56.1	12.3
Wives	15.2	35.7	12.6	5.9
Other relatives	15.5	42.1	11.6	5.8
Unrelated relatives	17.6	13.3	19.6	18.0
Males	56.9	22.9	56.5	9.9
Females	43.1	21.7	43.5	11.6
White	78.8	25.0	76.1	9.1
Blacks	19.4	12.5	21.9	23.2
Metropolitan residents	63.5	22.2	63.7	9.7
Nonmetropolitan residents	36.5	22.3	36.3	12.6

Source: Tabulations based on Current Population Survey data.

of the differences--those between blacks and whites and those between metropolitan and nonmetropolitan residents--reflect straightforwardly the diversity of their employment problems. Other differentials--those between the sexes and between persons with differing family status--are in part definitional since only household heads are included in the low earnings categories.

The EEI for blacks in March 1974 was 2.6 times that for whites, or more than the 2.1 ratio of adjusted unemployment rates. The unemployment rate clearly understates the disparity in hardship. Only an eighth of unemployed whites were poor, compared to a third of blacks; two-fifths of the former were in households with above-average income, compared to a fifth of the latter. Blacks were more frequently low earners and discouraged workers. In all categories, they were less likely to be in households with above-average incomes. A fourth of the subemployed whites were screened out by the upper income standard, compared to an eighth of blacks. Yet the average household income of blacks with inadequate employment and earnings was a tenth less than that of whites; half of the blacks compared to less than two-fifths of whites were living in poverty.

Whatever their relative position, there is no doubt about the severity of blacks' employment and earnings. Among black female family heads, the EEI was a staggering 56.0 percent and among unrelated females 32.5 percent. With such limited chances of success in the labor market, it is easy to understand why many find welfare an acceptable option.

According to the EEI, inadequacy is a sixth higher among female than male labor force participants. The difference would be greater if wives were included in the low-earnings categories. Two-fifths of women heading families had inadequate employment and earnings, more than four times the rate among male heads. Women in the adjusted labor force were 36 percent more likely to be among the unemployed in the EEI, 64 percent more likely to be employed part-time involuntarily, and 70 percent more likely to be discouraged and in a household with below-average income.

The EEI yields a different picture of the spatial distribution of hardship than the unemployment rate. In March 1974, 70 percent of the jobless resided in metropolitan areas compared with 64 percent of persons with inadequate employment and earnings. While the unemployment rate was virtually identical inside and outside SMSAs, the EEI in metropolitan areas was 2.9 percentage points less than in nonmetropolitan areas where low earnings were much more prevalent.

The EEI, 1968-1975

The EEI index has been calculated for March 1968 through 1975. This was a particularly turbulent period. It opened with a tight labor market which had attracted numerous secondary workers into the labor force. The major concern was with structural problems remaining after a lengthy boom. Social expenditures were rising rapidly and welfare had become a major political issue. Inflation was intensifying, as prices followed the Phillip's curve pattern. At the end of 1969, a decline in Vietnam war spending and some domestic belt-tightening to control inflation led to a substantial rise in unemployment. Recovery began in 1972 and became vigorous in 1973. Whether because the rebound was too rapid or because of exogeneous factors, inflation shot upwards and unemployment did likewise. By March 1975, forced idleness had reached massive proportions. What light does the EEI shed on these economic fluctuations?

Patterns of Change

There are several considerations in using the index to assess secular and cyclical changes. The upper adequacy screen, based on mean income, rises over time with inflation and any real gains, while the poverty threshold used as a lower screen in the earnings categories is adjusted only for changes in the cost of living.

Over a lengthy period, the relative well-being of persons screened in and screened out will change depending on the rate of increase in real income (which was slight between 1968 and 1975). The definition excludes persons 65 years old and over and students age 16 to 21 years as well as wives and other family members from the low earning categories. Secular changes in the composition of the work force may, therefore, influence the EEI (just as they affect the meaning of the unemployment rate).

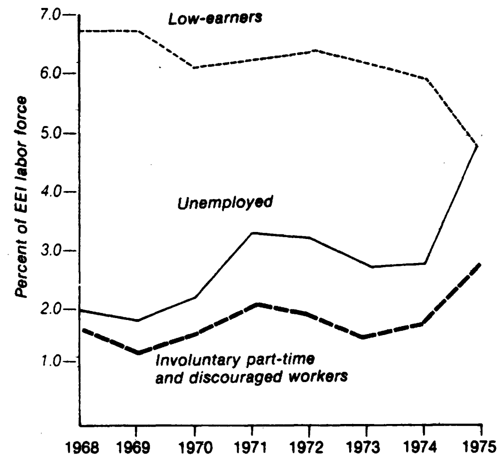
In interpreting cyclical changes, it is important to remember that the EEI's employment-related components--unemployment, discouragement, and involuntary part-time employment--are based on the current status in the survey week, while the earnings-related components are based on income over the preceding 12 months. Labor market changes are picked up immediately in the employment-related segments, but the earnings impacts lag.

Despite these complications, the EEI provides some useful perspectives on labor market developments between 1968 and 1975. As expected, the pervasiveness of economic hardship is affected by business conditions and fluctuates with the level of unemployment. The unemployment rate declined between 1968 and 1969, rose the next two years, and fell somewhat between 1971 and 1973. It levelled off between 1973 and 1974, then increased dramatically in 1975. The directions of change in the EEI were precisely the same (Table 5).

The important difference was that the EEI index fluctuated less than the unemployment rate. In the recessionary 1970s, the EEI index rose only a fourth above the 1968 level compared to a 140 percent increase in the CPS-reported unemployment rate. The unemployment rate rose 31 percent between 1969 and 1970, while the EEI index increased only 2 percent. In the subsequent recovery, the rate of joblessness declined 15 percent between 1972 and 1973, while the EEI index went down by 9 percent.

Two major factors explain these differences. Unemployment is only one segment of the hardship total (27 percent in 1974). Any percentage change in this component alters the EEI by a lesser percentage. The other factor is because the unemployed tend to be drawn from low-earning workers who may otherwise be counted in the EEI. Many intermittently employed may be included in the EEI because of low earnings even before they are forced into complete idleness (Chart 1). This offsets

Chart 1
EEI components as percent of adjusted labor force, 1968-1975



the fact that the number of discouraged and involuntary part-time workers also tend to increase with unemployment. Hardship is not just a cyclical problem. Though worsened by recession, it exists in serious dimensions even in what we have come to consider the best of times.

The Recession's Impact

The 1975 recession was the severest economic dislocation since World War II. The number of unemployed rose from a seasonally

Table 5. Unemployment and the EEI, 1968-1975

	1968	1969	1970	1971	1972	1973	1974	1975	Year-to-year changes 1968-1969	1969-1970	1970-1971	1971-1972	1972-1973	1973-1974	1974-1975
Persons with inadequate employment and earnings (thousands)	8,099	7,752	8,184	9,647	9,942	9,189	9,513	12,196	-347	432	1,463	295	-753	324	2,683
CPS unemployed (thousands)	2,929	2,746	3,733	5,175	5,215	4,512	4,755	8,359	-183	987	1,442	40	-703	243	3,604
CPS unemployment rate (percent)	3.8	3.5	4.6	6.3	6.1	5.2	5.3	9.1	-8	31	37	-3	-15	2	72
EEI index (percent)	10.4	9.8	10.0	11.6	11.5	10.5	10.5	13.2	-6	2	16	-1	-9	0	26

Source: Tabulations based on Current Population Survey data.

adjusted total of 2. 7 million in December 1968 to 4. 2 million in October 1973 after the relatively mild setbacks at the start of the decade. Unemployment then peaked at 8. 3 million in May 1975. Yet, there was surprisingly little public clamor over this slump--no riots, no large-scale marches on Washington, not even much rhetoric. The Republican administration continued to claim that inflation was the number one enemy, while Democrats with an overwhelming majority in Congress introduced only modest countercyclical programs while failing to override the vetoed spending measures. What was the reason for this quiescence?

The EEI offers one explanation. This index suggests that economic hardship did not increase as sharply as unemployment. In 1969 the EEI stood at 9. 8 percent and it was 10. 5 percent in 1974, after recovery from the short recession. In 1975 the EEI rose to 13. 5 percent. But if inadequacy had risen proportionately with joblessness, more than a fourth of the labor force would have faced economic hardship in March 1975.

As indicated, rising unemployment has a somewhat delayed effect because the intermittent employment category is based on the previous year's experience. Reflecting widespread joblessness in 1975, and the persisting high unemployment into 1976, the EEI total will probably rise further. Yet the 1975 index should give a good indication of the recession's impact. Unemployment had risen precipitously during the final quarter of 1974, from 5. 0 million in October to 6. 1 million in December, and to 8. 4 million by March 1975. The rise in unemployment at the end of the year had limited impact upon total earnings during the year.

Between March 1974 and 1975, the number of unemployed rose by 3. 5 million after excluding the elderly and students (Table 6). If the unemployed in households with above-average income are not counted, the increase was only 2. 1 million. The proportion of the unemployed screened out by the upper income standard remained constant at 36 percent in 1974 and 1975. In March 1975, 62 percent of the 2. 7 million excluded unemployed were wives or other relatives, and 10 percent were unrelated individuals.

Table 6. Changes in subemployment and inadequacy, March 1974-1975 (thousand persons)

	Subemployed 1974	Subemployed 1975	Percent Increase	EEI 1974	EEI 1975	Percent Increase
Total	12,250	17,113	39.7	9,513	12,196	28.2
Unemployed	3,889	7,343	88.8	2,518	4,645	84.5
Discouraged	585	1,153	97.1	452	826	82.7
Low-paid fully employed	2,076	2,121	2.2	1,897	1,899	---
Intermittently employed	3,702	3,270	-11.7	3,462	3,052	-11.8
Involuntary part-time workers	1,908	3,226	69.1	1,184	1,774	49.8
Percent of adjusted labor force	13.6%	18.5%		10.5%	13.2%	

Source: Tabulations based on Current Population Survey data.

While many victims of recession did not have serious needs, the conditions of others with already inadequate employment and earnings situations deteriorated even more. Workers with intermittent employment in the previous year fell 290, 000 between 1974 and 1975. The number of low-paid fully employed household heads did not change noticeably. This pattern was somewhat different than in the previous recession when the number of fully employed heads declined precipitously while the intermittently employed family heads increased. It might be surmised that the extended slack labor market had already trimmed the ranks of the low-paid workers in stable jobs, and the victims of the major slump were those who had already been affected by intermittent idleness.

The number of discouraged workers in households with below-average income rose by four-fifths between 1974 and 1975 to a level six times that in 1969. Many workers experienced shortened work-weeks, and the number employed part time involuntarily increased by half to double the 1969 level.

Overall, then, the total with inadequate employment and earnings increased by 2. 7 million, a fifth less than the increase in the number of unemployed. The proportion of the adjusted labor force with inadequate employment and earnings rose only by a fourth, compared to the 72 percent rise in the unemployment rate. Even in the most severe business downturn since the Great Depression, the continuing structural problem of hardship far outweighed the cyclical impacts. While headlines focused on the rise in unemployment, the increase in deprivation due to low earnings was much less and this may explain the limited social unrest generated by the economic downturn. The corollary, of course, is that when unemployment recedes it should not be assumed that the real problems have been eliminated.

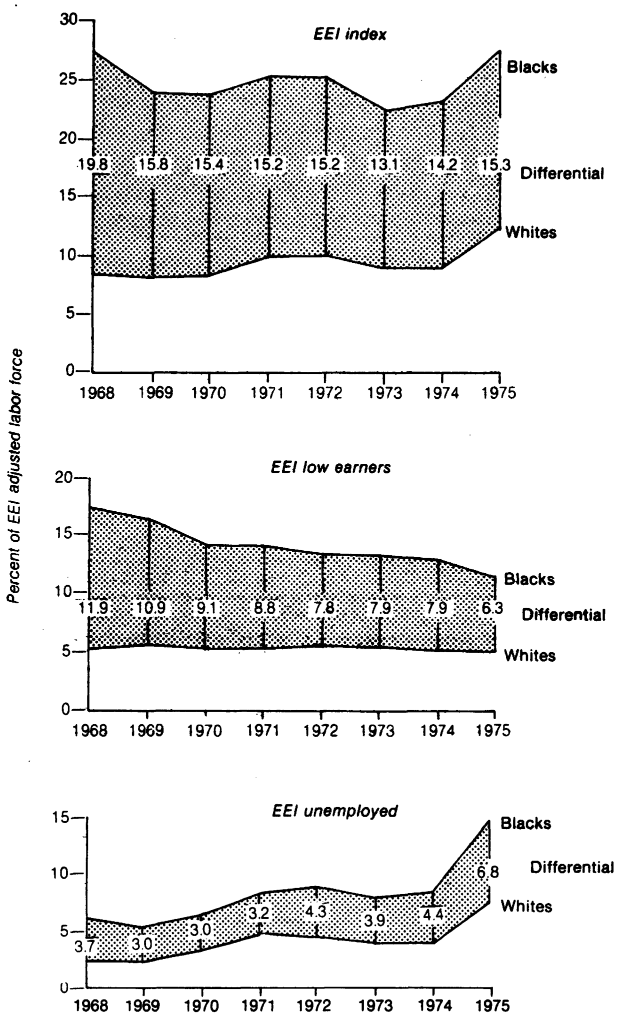
A Perspective on Racial Progress

In the 1960s there was a concerted effort to improve the employment status of minorities through manpower programs and equal employment opportunity action. The tight labor market provided a conducive climate, since those at the end of the labor queue tend to move up relatively, as well as absolutely, in good times. In the 1970s the government's commitment slackened, or, at least, its rhetoric favored a policy of benign neglect. The gainers in the tight labor markets become the losers in the recession. What, then, has been the end result for minorities?

The official unemployment statistics tell a not too pleasant story. Joblessness declined in 1969 to 3. 1 percent for whites and 6. 4 percent for nonwhites. The respective rates deteriorated to 7.8 and 13.9 percent in 1975. The nonwhite/white unemployment ratio fell from 2.1 to 1.8, but the unemployment rate differential increased from 3.3 to 6.1 percentage points. If nonwhites had done as well as whites in 1969, 295,000 more would have been employed in 1969 and 553,000 more in 1975.

The victims of unemployment and low earnings even in prosperous times have little to loose in economic slumps. Accordingly the EEI shows no further deterioration in the conditions of blacks between 1968 and 1975. The index for blacks remained virtually unchanged at 28 percent compared with a sharp increase from 8.5 to 12.5 percent for whites. The black/white inadequacy ratio declined from 3.2 to 2.2, while the gap was reduced from 19.8 to 15.3 percentage points. Before the severe economic setbacks of 1974, inadequacy was falling quite rapidly among blacks, reaching a low of 22.2 percent in 1973, compared to the upward drift of the white rate (Chart 2).

Chart 2
Incidence of low earnings, unemployment, and and inadequacy for blacks and whites, 1968-1975



The relative improvement for blacks is related to a decline in low earnings. The percentage of blacks in the adjusted labor force earning less than a poverty level wage fell from 17.3 percent in 1968 to 13.0 percent in 1974, while for whites the proportion declined only from 5.4 to 5.1 percent. The proportion for blacks went down even further to 11.4 percent in 1975, but this was probably due to increased joblessness among the otherwise intermittently employed.

Some blacks who would have had inadequate employment and earnings abandoned the labor force in preference to income support or other activities not counted as work by the Current Population Survey. As defined by the EEI index, the black male participation rate fell from 77.3 percent in 1968 to 72.6 percent in 1974, while the white male rate fell by .8 percentage points to 78.5 percent. If the decline for blacks had been the same as for whites, and the differential had all been added to the ranks of those with employment problems, the black EEI would have been 26.3 rather than 23.3 percent in 1974. This is an extreme assumption, however, and it does not deny that those leaving the labor force (including, for instance, many males receiving disability insurance or early retirement) were better off than in low-paying jobs. On the whole, then, it would appear that despite the lack of aggressive public efforts in the 1970s, some absolute and relative progress was made before the massive recession.

An Economic Hardship Measure

The Employment and Earnings Inadequacy index demonstrates the feasibility and usefulness of a measure which considers the impact of unemployment, discouragement, low-earnings, involuntary part-time and intermittent work on household well-being. The EEI is an exploratory formulation; the controlling constraint was the need to base calculations on currently available data. The underlying relationships are likely to persist, however, in any reasonable measure of labor market-related economic hardship.

The Utility of the EEI

In normal times the jobless are a minority of those who might reasonably be considered in need. On the other hand, many unemployed do not face serious economic hardships. Indeed some are very well off. The EEI clearly demonstrates that there are many employed persons who do not rise above the poverty threshold, even if they work at full capacity. The unemployment rate is, therefore, a poor hardship measure. Substituting EEI concepts for traditional unemployment data suggests greater concentration of need outside metropolitan areas. The EEI also highlights the still dismal labor market prospects of blacks, especially black female family heads.

The data for 1968 through 1975 reveal consistent patterns of relationship between the components of the EEI. When joblessness rises, many low earners and intermittent workers are the victims. On the other hand, many of the additional unemployed are screened out by the upper adequacy standard; these are mostly secondary workers in households with above-average incomes. Hence, the inadequacy index fluctuates much less than the unemployment rate. Even during the worst recession since the 1930s, the EEI rose modestly, perhaps explaining why the "social dynamite" that might have accompanied widespread joblessness never exploded. There is some evidence that the gap between blacks and whites closed between 1968 and 1975, though this was achieved by white setbacks rather than black gains over the period.

In brief, the EEI provides a reasonable and comprehensible assessment of needs. It seems to be a consistent measure, making sense when analyzed over time. Most importantly, it provides significant insights into labor market realities. Why, then, has no such measure been refined and officially tabulated to supplement the unemployment rate?

Inexcusable Procrastination

Based on earlier works by Secretary of Labor Willard Wirtz in 1967 and 1968, by the staff of the Senate Subcommittee on Employment, Manpower, and Poverty in 1972, and previous EEI calculations, Congress recognized the value of a needs index and directed the Department of Labor to "develop preliminary data for an annual statistical measure of labor market related economic hardship in the nation" (Section 312 (c) of the 1973 Comprehensive Employment and Training Act). In the 1975 Manpower Report of the President, the Labor Department reported on its progress: "considerable conceptual work must be done in the development of statistics on economic hardship. When satisfactory definitions and criteria have been developed, ways to use these in analyzing economic hardship and underemployment can be examined (p. 189)." Translating this bureaucratic jargon: no data had been collected and no new definitions tested. This remains the case today.

It is difficult to rationalize the failure of BLS to carry out the clear congressional mandate. The "conceptual work" in developing, analyzing and presenting the EEI for 1968-1975 amounted to less than one-half a man-year. A number of improvements and alternatives have been proposed which could be tested with little effort. The cost of developing the computor program, calculating the index for the eight years, and running several validation tests was less than $10,000. The incremental cost of calculating the index for any given year is $500. The shortage of conceptual or financial resources in the Department of Labor is clearly not the real constraint.

The lack of progress simply reflects a lack of priority. The administration was understandably reluctant to admit that conditions might be worse than already staggering unemployment rates suggested (although, paradoxically, the index would have demonstrated that conditions did not deteriorate as severely in 1975 as unemployment tallies indicated). The massive increase in joblessness diverted attention to other issues. But even more basically, administration economic policy shapers were apparently opposed to the underlying concepts of a needs index which would focus attention on deep-seated structural economic problems.

Can We Ignore Hardship?

The EEI and other economic hardship measures are based on the notions that: (1) the inadequacy of earnings is as important as the availability of employment; (2) unemployment and earnings problems are interrelated and compounded for a significant minority of all workers; (3) the gravity of employment problems is primarily related to their impact on household income; and (4) those with the most severe problems are the ones who should be given attention. In contrast, the prevailing view of the many policymakers in the first half of the 1970s seemed to be that any job was better than no job, that low earnings due to intermittent work was a reflection of limited work commitment, and that earnings provided in the labor market could somehow be divorced from family income needs.

These arguments which tried to explain away hardship sound disturbingly like the pre-depression neoclassical theories which dismissed mass unemployment as a transitional phenomenon. Problems do not disappear simply because we refuse to recognize them. Just as a new set of statistics were introduced in the late 1930s to measure unemployment, it is necessary to overhaul and supplement current economic indicators. The unemployment data are not an adequate measure of economic hardship. The need is to attack the problems of wage adequacy and intermittent work by focusing public efforts on work force participants who face economic deprivation in good times as well as in recessions. A first step is to develop and refine measures of labor market-related economic hardship.

INFLUENCE OF THE POVERTY LINE ON FEDERAL PROGRAMS

George Grob
Department of Health, Education and Welfare

There are many poverty measures used by the federal government. One of them, which is variously referred to as the Orshansky, Social Security, "OMB", or Census poverty measure, holds a place of pre-eminence. In one sense, it is more official than the others. It is certainly more widely used. This measure was originally developed by Mollie Orshansky of the Social Security Administration in 1964 and was, with revisions, officially adopted in 1969 by the Office of Management and Budget (OMB) as the Federal Government's official statistical measure of poverty. The measure is built around the Department of Agriculture's economy food plan of 1961 and the national average ratio of family food expenditures to total family after-tax income as measured in the 1955 Household Food Consumption Survey. It consists of 124 separate poverty cutoffs differentiating families by size, number of children, age and sex of head and farm or nonfarm residence. The cutoffs are updated annually by changes in the Consumer Price Index. The poverty cutoff for a male headed nonfarm family of four with two children in 1974 was $5000. According to the Census Bureau's report based on the March 1975 Current Population Survey, in 1974 there were approximately 24.3 million persons, or 12 percent of the population, poor by this definition. There were 5.1 million poor families, 9.2 percent of all families.

The 124 official poverty lines are somewhat awkward to use for many applications. A simplified list of weighted average poverty lines, or "cutoffs", is published annually by the Census Bureau; the weighted average cutoff for 1974 are shown in table 1.
Whatever their original purposes were, either in their design or in their official adoption as a federal statistical tool, the Orshansky poverty lines have been adapted for many uses and they have had pronounced effects on individuals, geographic areas, and social institutions.

In order to gauge the effects of changing these poverty lines, it is useful to distinguish three uses to which they are put. First, the poverty lines, along with annual reports form the Census Bureau on the number and characteristics of the poor, serve as a measure of the nation's progress in reducing the extent of poverty. Second, they are used as a statistical tool to identify needy populations for the purpose of designing or evaluating federal programs which are aimed at assisting the poor. Thirdly, they are used for administrative purposes to dispense funds to individuals or areas in need.

Poverty Lines as Measures of Progress

The poverty lines remind us of the existence of the poor and needy. Their effect on the national consciousiness and consequently on national initiatives on behalf of the poor is very real. A concern is sometimes expressed that the official poverty counts have consistently decreased and will probably continue to do so. It is feared that this will lead to an artificial solution of the problem of poverty inasmuch as poverty will only seem to go away through the device of federal statistics. Proposals to revise the current poverty measure by using more current statistics or to change the basis of poverty measurement by using relative concepts based on on median family income sometimes stem from this appreciation of the far reaching effects of the poverty. Of course, such proposals also stem from alternative economic concepts, from a desire to use current data, and from practical considerations like availability of raw data and the suitability of these measures for various administrative purposes. Furthermore, some argue in favor of the lower poverty counts because they believe that there are in fact fewer poor people than in years past. No matter what the underlying motivations, these poverty measures would tell different stories to the public over long periods of time and therefore would probably affect the way people think about the poor and about the effectiveness of government programs for the poor.

According to the data from the Current Population Survey, under the official poverty measurement system (when backdated by the Consumer Price Index), the number of poor families was reduced from 18.5 percent in 1959 to 9.2 percent during 1974. Revising the official poverty line on the basis of current nutrition standards, food plans, food prices, and a higher multiplier reflecting more recent overall consumption data would raise the poverty lines in real terms and lessen the amount of progress shown in reducing the extent of poverty. A relative poverty line based on 50 percent of national median family income would consistently show about 19 percent of all families as poor over the past fifteen years, although at ever higher real income levels.

Poverty Lines As a Statistical Tool

The largest social insurance programs like Social Security, Railroad Retirement, and Veterans benefits; the public assistant programs like Aid to Families with Dependent Children and Supplemental Security Benefits; and in kind programs like food stamps, Medicare, and Medicaid generally do not incorporate the official poverty lines either for setting their income eligiblity levels or their benefit levels. Now, there are many reasons why there is not a direct link to the poverty lines. Imposing a cutoff of benefits at the poverty line could produce a sudden decrease in benefits as a person's income

Title I of the Elementary and Secondary Education Act, as amended, is an example of a program using a poverty measure as part of an allocative formula. Approximately $1.5 billion are distributed annually partly on the basis of the number of poor school-age children in each county. Where county boundaries are not coterminous with school districts, county amounts are then suballocated to the school districts by state departments of education. Children living in school attendance areas which have an incidence of poverty as high as or higher than the district-wide average are eligible for Title I services. Children are selected for participation on the basis of educational deficiencies, regardless of family income. Other programs which distribute funds to needy areas are Title I of the Housing and Community Development Act of 1974, and grants made to Community Mental Health Centers with disproportionate poverty populations in health catchment areas.

The Community Services Administration (CSA), formerly the Office of Economic Opportunity, provides an example of administrative use of the Orshansky measure for determining individual income eligibility. Uniform income eligibility standards are issued for poverty-related programs administered by CSA such as the Community Action Agencies. Based directly on the official Federal poverty measure, these guidelines eliminate many of the distinctions and smooth some of the remaining variations. The CSA poverty thresholds allow for variations by family size, with equal dollar increments or additional family members. Like the Orshansky poverty matrix, cutoffs for farm families are 15 percent below those for nonfarm families. In addition, two major geographic variations are partly taken into account by raising the cutoffs by 25 percent for Alaska and by 15 percent for Hawaii. The CSA thresholds are updated annually according to the change in the Consumer Price Index. Table 2 presents the 1974 CSA poverty cutoffs for the continental United States, along with the comparable weighted average cutoffs from the Orshansky system from which these were derived. For each of the weighted Orshansky averages in Table 2, the numbers in parentheses indicate the range of variation across all other poverty lines for the same family size in the full Orshansky matrix of 124 poverty lines.

Similar guidelines are issued by the Secretary of Labor and the Secretary of Agriculture for their programs for low income persons. Examples of such programs are the Comprehensive Employment and Training Act and the Child Nutrition Programs.

Table 2 1974 Orshansky Poverty Thresholds and Community Services Administration Income Eligibilty Guidelines

Family Size	Community Services Administration		Orshansky Poverty Measure	
	Nonfarm[a]	Farm[a]	Nonfarm	Farm
1 person	$2590	$2200	$2495(2358-2659)[b]	$2092(2004-2260)
2 persons	3410	2900	3211(2948-3724)	2707(2506-3165)
3 persons	4230	3600	3936(3568-4223)	3331(3033-3590)
4 persons	5050	4300	5038(4900-5252)	4302(4165-4465)
5 persons	5870	5000	5950(5781-6232)	5057(4914-5298)
6 persons	6690	5700	6699(6457-7087)	5700(5489-6024)

SOURCE: Federal Register, Vol. 40, No. 132 (July 9, 1975), p. 28794. U.S. Bureau of the Census, "Characteristics of the Population Below the Poverty Level: 1974," Current Population Reports, Series P-60, No. 102, Table A-2.

[a] All States except Alaska and Hawaii. The thresholds for Alaska are 25 percent higher, and those for Hawaii are 15 percent higher.

[b] Figures in parentheses indicate the range of variation across the thresholds in the measure.

A single legislative act may contain both an allocative formula for distributing fixed program funds and eligibility criteria for determining which individuals in each area are entitled to receive assistance. Furthermore, entirely different poverty measures may be used in the various stages of one program, as with the Comprehensive Employment and Training Act (CETA). The measure of poverty used for distributing CETA funds is a single-dollar threshold ($7,000 in 1969 dollars), with the allocation based partly on the number of families in an area with an income below that level. At the local level, however, individual eligibility is determined in part on poverty guidelines like those just described for the CSA and in part on other factors like unemployment or underemployment.

In some programs the Orshansky thresholds are used but in a modified form. Some of the benefits of the Child Nutrition Program, for example, are available to families with incomes up to 195% of the poverty lines. Another common adaptation is to include status as a recipient of welfare, in conjunction with the poverty thresholds, to identify target populations. A person may be eligible for food stamps for example either by the fact that his family income falls below the poverty cutoff or because he is in a family receiving welfare payments. A special criterion relating family income to the cost of the food is also used in this program.

Except for the Child Nutrition Program and possibily the food stamp program, the budgetary impact of a change in the official poverty matrix would not be great for all the programs just named. In most cases, the poverty thresholds are used to distribute equal fixed appropriated funds to States, counties, cities, or other areas on a formula basis or to needy persons on a first-come, first-served basis or on a most-in-need basis.

Raising the poverty lines would not cause federal appropriations to rise in the near term. Use of the poverty lines for administrative purposes affects how thinly and to whom the limited funds are distributed. Thus, the relevant factor in determining who will benefit from raising or lowering the poverty lines is not the poverty rate or total poverty count but rather the change in the distribution of the poor among the various demographic subgroups of the population.

The two concepts, poverty rates and distribution can be compared in the following two tables. The official poverty lines were lowered to 75% of their official value, then raised to 125%, 150%, and 200%. These were alternately used as new "poverty" lines to measure the demographic characteristics of the poor at higher or lower poverty levels. From these tables, particularly Table 4, it is possibile to identify whites, working poor, male headed families, and the elderly as groups whose representation in the poverty population would increase with higher poverty lines, and who would thus be the primary beneficianies of higher poverty lines.

Table 3

Poverty Rates by Selected Characteristics for Alternative Poverty Levels, 1974

Characteristics	Universe (in thousands)	Current Measure	Scaling of Current Measure			
			75%	125%	150%	200%
Persons						
Total	209,343	11.6%	6.9%	16.5%	21.6%	33.1%
Living arrangements:						
In families	190,471	10.2	6.1	14.6	19.4	31.0
Male headed	167,227	6.5	3.6	10.2	14.7	26.4
Female headed	23,245	36.8	24.2	46.0	53.1	64.3
Unrelated individuals	18,872	25.5	15.1	36.2	43.9	54.8
Race:						
White	182,355	8.9	5.2	13.2	17.9	29.2
Black	23,704	31.5	20.1	41.5	49.2	62.4
Other	3,284	15.1	9.5	21.7	26.0	38.8
Age:						
Under 5 years	16,002	16.7	10.9	22.8	29.3	44.1
5-17 years	49,800	15.1	9.2	20.4	25.9	39.1
18-64 years[a]	122,414	8.8	5.5	12.5	16.5	26.3
65 years & over	21,127	15.7	6.7	25.9	35.1	50.3
Families						
Total	55,712	9.2	5.5	13.3	17.9	28.8
Sex of head:						
Male	48,470	5.7	3.1	9.2	13.4	24.2
Female	7,242	32.5	21.2	40.8	47.6	59.5
Presence of children:						
None	24,381	5.1	2.5	8.7	12.7	22.1
One or more	31,331	12.4	7.8	16.9	21.9	34.0
Employment status of head:						
Employed	40,419	5.1	3.0	7.8	11.1	20.4
Unemployed	2,797	16.1	10.9	22.8	29.2	43.0
Not in civilian labor force	12,497	20.9	12.3	29.1	37.1	52.7
Source of income:[b]						
Earnings	49,529	6.4	3.8	9.8	13.7	23.8
Social Security	12,162	10.0	4.5	16.9	24.3	39.9
Public assistance	4,359	46.9	27.4	59.4	68.3	79.8
Unrelated Individuals						
Total	18,872	25.5	15.1	36.2	43.9	54.8
Sex:						
Male	7,890	20.4	13.2	28.0	34.0	44.5
Female	10,981	29.3	16.4	42.1	51.0	62.1
Employment status:						
Employed	9,660	13.2	9.0	18.7	23.3	33.6
Unemployed	930	32.4	23.7	40.2	45.3	59.4
Not in civilian labor force	8,282	39.2	21.1	56.2	67.7	79.0
Source of income:[b]						
Earnings	11,609	14.6	9.2	20.9	26.0	37.1
Social Security	6,982	30.7	11.9	49.6	62.8	75.7
Public assistance	1,656	62.7	30.6	79.8	90.8	96.1

Table 4

Distribution of the Poverty Population by Selected Characteristics for Alternative Poverty Levels, 1974

Characteristics	Total Population	Current Measure	Scaling of Current Measure			
			75%	125%	150%	200%
Persons						
Number (in thousands)	209,343	24,260	14,538	34,615	45,211	69,389
Living arrangements:						
In families	91.0%	80.1%	80.5%	80.3%	81.7%	85.1%
Male headed	79.9	44.8	41.7	49.4	54.4	63.6
Female headed	11.1	35.3	38.8	30.9	27.3	21.5
Unrelated individuals	9.0	19.9	19.5	19.7	18.3	14.9
Race:						
White	87.1	67.2	65.1	69.5	72.3	76.9
Black	11.3	30.7	32.7	28.4	25.8	21.3
Other	1.6	2.0	2.1	2.1	1.9	1.8
Age:						
Under 5 years	7.6	11.0	12.0	10.5	10.4	10.2
5-17 years	23.8	31.0	31.7	29.3	28.5	28.0
18-64 years[a]	58.5	44.3	46.5	44.4	44.7	46.5
65 years and over	10.1	13.6	9.8	15.8	16.4	15.3
Families						
Number (in thousands)	55,712	5,109	3,052	7,437	9.948	16,036
Sex of head:						
Male	87.0	54.0	49.8	60.3	65.4	73.2
Female	13.0	46.0	50.2	39.7	34.6	26.9
Presence of children:						
None	43.8	24.2	20.0	28.6	31.2	33.6
One or more	56.2	75.8	80.0	71.4	68.8	66.4
Employment status of head:						
Employed	72.5	40.1	39.5	42.4	45.2	51.4
Unemployed	5.0	8.8	10.0	8.6	8.2	7.5
Not in civilian labor force	22.4	51.1	50.5	50.0	46.6	41.1
Source of income:[b]						
Earnings	88.9	62.1	61.5	65.0	68.0	73.4
Social Security	21.8	23.9	17.9	27.6	29.7	30.2
Public assistance	7.8	40.0	39.2	34.8	29.9	21.9
Unrelated Individuals						
Number (in thousands)	18,872	4,820	2,841	6,832	8,284	10,333
Sex:						
Male	41.8	33.3	36.7	32.3	32.3	34.0
Female	58.2	66.6	63.3	67.7	67.6	66.0
Employment status:						
Employed	51.2	26.4	30.7	26.5	27.2	31.4
Unemployed	4.9	6.3	7.7	5.5	5.1	5.3
Not in civilian labor force	43.9	67.3	61.5	68.1	67.7	63.3
Source of income:[b]						
Earnings	61.5	35.1	37.6	35.5	36.5	41.7
Social Security	37.0	44.5	29.3	50.7	53.0	51.1
Public assistance	8.8	21.6	17.8	19.4	18.2	15.4

SOURCE: Special tabulations by the Census Bureau from the March 1975 Current Population Survey.

[a] Includes a small number of heads, wives, and unrelated individuals 14-17 years of age.

[b] A family or unrelated individual may have more than one source of income. Consequently, the totals for these categories exceed the numbers of families and unrelated individuals.

Table 1.

Weighted Average Poverty Cutoffs in 1974 by Size of Family and Sex of Head, by Farm-Nonfarm Residence

Size of Family Unit	Total	Nonfarm			Farm		
		Total	Male Head	Female Head	Total	Male Head	Female Head
1 Person (Unrelated Individual)	$2,487	$2,495	$2,610	$2,413	$2,092	$2,158	$2,029
14 to 64 Years	2,557	2,562	2,658	2,458	2,197	2,258	2,089
65 Years and Over	2,352	2,364	2,387	2,357	2,013	2,030	2,002
2 Persons	3,191	3,211	3,220	3,167	2,707	2,711	2,632
Head 14 to 64 Years	3,294	3,312	3,329	3,230	2,819	2,824	2,706
Head 65 Years and Over	2,958	2,982	2,984	2,966	2,535	2,535	2,533
3 Persons	3,910	3,936	3,957	3,822	3,331	3,345	3,133
4 Persons	5,008	5,038	5,040	5,014	4,302	4,303	4,262
5 Persons	5,912	5,950	5,957	5,882	5,057	5,057	5,072
6 Persons	6,651	6,699	6,706	6,642	5,700	5,700	5,702
7 or More Persons[a]	8,165	8,253	8,278	8,079	7,018	7,017	7,066

SOURCE: U.S. Bureau of the Census, "Characteristics of the Population Below the Poverty Level: 1974," Current Population Reports, Series P-60, No. 102, Table A-2.

[a] Represents an average for families with 7 or more persons.

reaches the poverty line. This may cause serious inequities or adversely affect work incentives. Also, the purpose of a program may not be congruent with the Federal statistical definition of poverty. Moreover, some believe that it would be too expensive to guarantee a poverty level income to all persons through these programs.

The important factor concerning the relationship between the poverty lines and these programs is not the absence of a direct link between the two; it is the general surprise which is usually registered when someone is first informed about the lack of the connection that is important. Even after going through the arguments just summarized about why there need not be a direct link, there is a persistant and nagging feeling that there ought to be one. This notion is quite strong. It is shared by many program analysts, public leaders, and the public. The author has witnessed delibrations on subjects like welfare reform, health financing and other social programs and has observed first hand how participants both in formal meetings and at informal discussions instinctively reach for the poverty line or some adaptation of it as a rough measure of where they want to come out. Much fine tuning of program parameters takes place after the basic program structures have been roughly estimated, but the first cut is often made by reference to the official poverty lines. Congressional and executive branch staff members have asked for statistics to help formulate program benefit levels at "somewhere around 150% of the poverty lines" or "just right above the poverty line". If the poverty lines were higher or lower, then program analysts and legislators throughout the government would be thinking about programs with higher or lower benefit levels and someone would eventually be proposing them as serious candidates for review. A good example of this phenomenon can be seen in the fact that the guaranteed benefit levels of the Supplemental Security Income Program are just slightly below the poverty lines. Another example is the recent ongoing public debate on whether the non poor, specifically those with incomes above the Orshansky lines, should be eligible to purchase food stamps. Less obviously, federal income taxes are keyed to these poverty lines to the extent that an unofficial but nevertheless real policy has existed to set the personal exemptions and the standard deductions near the official poverty lines so that the poor won't have to pay federal income taxes.

Admittedly, this analysis is disappointing in as much as the effects, however real, are difficult to measure. Furthermore, the influence of the poverty lines on the level of the federal budget through the large social programs is a long range, sometimes permanently postponed, effect. In the near term, the federal budget would not increase with higher poverty lines. Rather, because of the way the lines are used for administrative purposes, limited federal funds would be shared by more and different people.

Poverty Lines As Administrative Devices

Some federal programs, either by law or by executive order, make direct administrative use of the official poverty lines. Federal programs for the poor differ in design. Some programs are designed to aid areas and some are designed to aid families or individuals directly. In the former case, the poverty measure is used in an allocative formula to distribute the appropriation, typically a fixed amount, among the subunits of the of the nation designated by the legislation. In the second type of programs, a poverty cutoff may be used as an income eligibility criterion for individual applicants.

For example, the current "poor" population is approximately 2/3 white and 1/3 non white. If the poverty lines were raised by 50 percent (so that the poverty line for a non farm family of four would be about $7500 in 1974) then the poor population would be almost 3/4 white. Similarly, under the current measure, 54 percent or slightly more than 1/2 of all poor families have male heads. If the poverty lines were increased by 50 percent, then 2/3 of the officially poor families would have male heads. If the poverty lines were doubled, then almost 3/4 of the poor families would have male heads.

As far as the geographic distribution of the poor is concerned the South would show a decreased share of the poor, and consequently of federal funds, while all other regions would show an increased share under higher poverty lines.

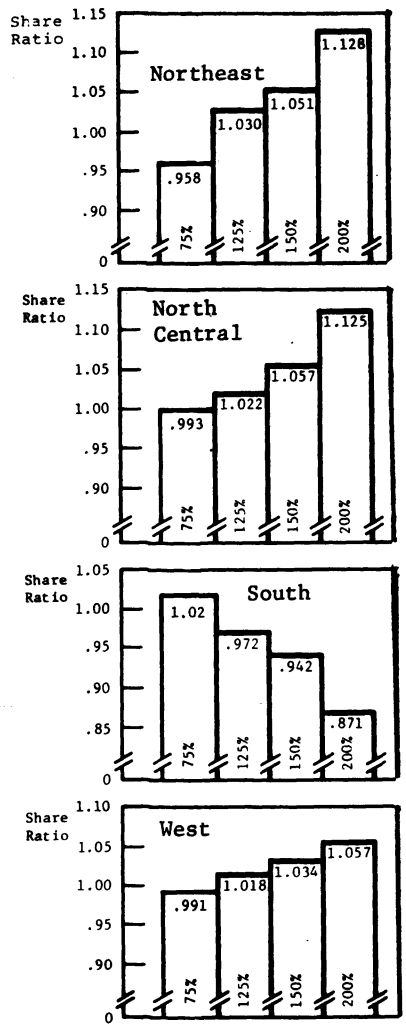
As an aid in comparing effects, the notion of a share ratio is used. The share ratio is defined as a regions percentage of the nation's poor population resulting from an alternative poverty level divided by its percentage resulting from the current level. Percentages are based on 1970 Decennial Census and reflect data for the year 1969. A ratio of 1.0 indicates that a region's share is unchanged by the alternative measure; a ratio smaller than 1.0 indicates a decreased share.

Figure 1 illustrates this. In each case, the bar on the graph represents the region's share of the poverty population under the various poverty levels relative to its share under the current level. For example, the Northeast contained 17.9 percent of the 27.4 million poor persons in the United States as counted in the Dencennial Census. Raising the current poverty thresholds by 25 percent results in a poverty population of 37.5 million persons, of which 18.4 percent resided in the Northeast. Therefore, the share ratio for the Northeast under the upward scaling of the poverty measure by 25 percent is 1.03 (18.4/17.9). If the poverty lines were increased by 50 percent, then the Northeast would increase its share of the nations poor by 5.1 percent, the North Central region by 5.7 percent, and the West by 3.4 percent. The South however, would decrease its share by 5.8 percent.

In addition to the regional patterns, the more populated states as a group increase their share of poor persons more than other states as the poverty lines are increased. When the poverty lines are set at 150 percent of the official thresholds, the national poverty rate is increased by from 13.9 percent to 24.3 percent. However, over half of this increase is attributable to low-income persons living in the eleven most populated states. As a group, these states contain 48 percent of the poverty population under the current poverty thresholds and 51 percent of the poverty population under poverty thresholds set at 150 percent of the official level.

The general principles just illustrated, that the federal budgets would not increase with

Figure I



Source: Special tabulations by the Census Bureau from the 1:100 Sample of the 1970 Census Population.

Note: The use of share ratios for the analysis of distributional effects of alternative poverty measures was introduced by Lawrence Brown, Office of the Assistant Secretary for Planning and Evaluation, Department of Health, Education, and Welfare. For further discussion of this, see Characteristics of Low-Income Population Under Alternative Poverty Definitions, Technical Paper XVIII to The Measure of Poverty, a report published by the Department of Health, Education and Welfare, April, 1976.

higher poverty lines but that limited funds would be shared by a larger and differently composed poor population, are not true in every case.

In some programs, eligibility is open-ended, in that the benefits are provided to all eligible persons requesting them; such programs would force the Federal budget upward if the poverty line were increased. The Child Nutrition program of the Department of Agriculture, for example, would be affected significantly by an increase in the poverty line. Based on poverty rates for children, as indicated in table 3, a 25 percent increase in the poverty line could result in an increase of more than 30 percent in this program's budget under current law.

A smaller effect would occur in the food stamp program because its income eligiblity criteria already generally exceed the poverty lines for most family sizes. There would be a much more significant effect on the food stamp program if it were redefined primarily on poverty criteria such as in recently proposed legislation and regulations. Obviously, the structure of the poverty guidelines would then be a most crucial program parameter.

Other Poverty Measures

Measures of poverty or income eligibility and concepts of need other than the Orshansky measure are also used in Federal programs. A single dollar threshold that is unchanged for family size is currently used as an eligibility criterion in the College Work-Study program authorized by the Higher Education Act of 1965. As mentioned earlier, this type of measure is also used in the allocation formula of the Comprehensive Employment and Training Act; one was also used in the allocation of funds under Title I of the Elementary and Secondary Education Act until 1974.

Another measure of income eligibility is one based on some percentage of median income. Title XX of the Social Security Act (social services) adopted 80 to 115 percent of median family income in each state as its standard. Title II of the Housing and Community Development Act of 1974 uses 50 and 80 percent of median family income in the "area" as it eligibility criteria. The use of standards based on median income is a relatively new development in Federal programs, although there is precedence in social service programs.

Finally, administrative and legislative references abound in terms that target Federal programs to the "disadvantaged, "needy," "dependent," "economically disadvantaged," and "individuals whose income and resources are insufficient." In many such references, the terms are employed without definition.

There is not sufficient room here to examine all viable poverty measurement techniques. Not all changes can be approximated by reference to higher or lower poverty lines. Some revisions, such as those based soley on substitution of new food plans for the 1961 economy plan, could raise poverty lines for some family types and lower them for others. For a fuller treatment, one should consult the report entitled The Measure of Poverty published in April 1976 by The Department of Health, Education and Welfare. There is also a series of technical papers connected with the report which bear on the subject of this paper.

This paper draws heavily, sometimes by verbatim transcription, from The Measure of Poverty. The intent was was to pull together those parts of the report which describe the generalized effects on beneficiaries of federal programs of changing the the official poverty lines. The author therefore must acknowledge, and is happy to do so, his gratitude to the members of the Poverty Studies Task Force, a federal interagency group which prepared the report, for use of the materials from the study. At the same time, it must be noted that this paper is an edited selection and organization of some of that material along with some original statements which were not subject to approval or even advance review by the Poverty Studies Task Force. Hence, incorrect interpretations of or wrong conclusions about the transcribed material are attributable to the author of this paper and not to the Poverty Studies Task Force.

To obtain copies of The Measure of Poverty or the technical papers, one may write to:

Office of the Assistant Secretary for Planning and Evaluation
Department of Health, Education and Welfare
200 Independence Avenue S.W.
Room 443 D, South Portal Building
Washington, D.C. 20201

REVISIONS OF STATISTICAL DEFINITIONS: WHAT IS A FARM?

R. J. Hildreth, Farm Foundation, and Gaylord Worden, ERS, USDA

The entity called a farm is central to our food and fiber data system. It is the central concept around which three-quarters of our food and fiber statistics are designed and collected. (2) Thus, the concept behind the statistic called "number of farms" and related derivatives such as the "net farm income per farm" are very important in developing a statisitical view of U.S. food and fiber production. Farm policies to adjust supplies of agricultural commodities and enhance or stabilize incomed earned from farming are significantly influenced by the statistical measures based on this concept.

In his recent book Social Information Processing and Statistical Systems -- Change and Reform, Edgar S. Dunn points out that there is a strong tendency to objectify the world; i.e. to conceive of it in terms of concrete objects. (3) One result of this is that we tend to treat entities as if they had "skin" in spite of the fact that an entity is an abstraction. This has proven to be the case in the attempt to change the definition of what is called a farm for the 1974 census of agriculture. We will describe some of this experience in trying to revise the statistical definition and attempt to state some broader implications for renewal of data systems.

Dunn also develops a taxonomy of names for entities. The first element he calls "identifiers." They are used to identify single entities, such as an individual farm. Secondly, the states and activities that characterize the entitities, he calls "descripters." The efforts for 1974 were to change both the identifiers and descripters of the entity called a farm.

The definition of a farm in use has been a place with sales of $250 or more of agricultural products or any place of 10 acres or more with sales of $50 or more. The proposed new definition is any establishment from which $1,000 or more of agricultural products is sold or would normally be sold. Thus, two aspects of the identifier are modified. One is to increase the minimum value of sales to be called a farm. The second is to change from the concept "place" to "establishment."

The changes in descripters are additions to the set of descripters used to classify data on farms. Additions are made to the number of classes of value of sales from the establishment. Another change implements the use of the more detailed type of product descripters contained in the four digit levels of the 1972 Standard Industrial Classification Manual. (4) Finally, a new major group of descripters is introduced that will provide the framework for separating farms into: (a) those farms which keep the operator primarily employed in farming (b) farms where the operator is only employed for a minority part of his work time, and (c) farms that are operated as a minority part of multi-establishment companies that have integrated or diversified into farming.

The definition of what is called a farm and related descripters have been changed a number of times since farms were first counted in the Agricultural Census of 1850 (table I).

Past changes, as well as the efforts for 1974 were made at the time of a Census of Agriculture which is taken each five years. The Secretary of Commerce has the authority to determine who will be included for each Census. The Department of Agriculture has generally worked closely with the Department of Commerce and made recommendations for change in census definition because they use the same identifier for reporting of other statistics about farms. A point should also be made that other program agencies of USDA and other government agencies such as the Internal Revenue Service and Social Security use program and reporting requirements that result in different definitions and thus different counts of the number of "farms."

The recommendations by the Department of Agriculture in 1973, in preparation for the 1974 census, were made after careful assessment of potential impact from the change. Consultation had been held with a number of advisory, Congressional, farm organizations, and professional groups. Several alternative changes were presented to these groups and their reactions to the alternatives were used in shaping the final recommendations.

The USDA wanted to make the change so that resulting statistics on farming would more nearly reflect the farming industry that existed today. A USDA spokesman at a Congressional hearing said the need was for "truth in Labeling" and development of statistics that were more meaningful for making public policy decisions. (5,7)

The USDA spokesman went on to explain that the increase of minimum sales level to $1,000 would mean that approximately 18 percent of old definition farms would be dropped from the statistics. Only about three-tenths of one percent of all farm products were sold from these units. These large number of very small farms were said to distort the commonly used measures for assessing the economic well being of farms, i.e. net income per farm. And these distorted statistics then, in turn, provide misleading information when decisions are made for farm price and income policies.

The USDA testimony stated that operators of these very small farms receive most of their family income from other sources (table II). They receive only very small benefits from farm price and income programs because of the small size of their farming activity, but they are not necessarily low income families. The people involved in these very small farming operations have needs just as do other nonfarm residents of rural America. These include jobs, quality education, health care and personal safety. But these needs are identified by statistics on the rural farm and rural

TABLE 1.-- Farm definitions used in Censuses of Agriculture

Year	Acreage limitations	Other criteria
1850 1860	None	$100 worth of agricultural products produced for home use or sale
1870 1880 1890	3 or more acres - less than 3 acres -	Any agricultural operations $500 worth of agricultural products sold
1900	None	Agricultural operations requiring continuous services of at least one person
1910 1920	3 or more acres - less than 3 acres	Any agricultural operations $250 worth of agricultural products produced for home use or sale; or constant services of at least one person
1925 1930 1935 1940	3 or more acres - less than 3 acres -	Any agricultural operations $250 worth of agricultural products produced for home use or sale
1945	3 or more acres -	Agricultural operations consisting of 3 or more acres of cropland or pastureland; or $150 worth of agricultural products produced for home use or sale
	less than 3 acres -	$250 worth of agricultural products produced for home use or sale
1950 1954	3 or more acres -	$150 worth of agricultural products produced for home use or sale
	less than 3 acres	$150 worth of agricultural products produced for sale
1959 1964 1969	10 or more acres - less than 10 acres -	$50 worth of agricultural products produced for sale $250 worth of agricultural products produced for sale
•1974 (proposed)	None	$1,000 or more worth of agricultural products produced for sale

TABLE 2.-- Off-farm income as a percent of realized net farm income, 1960-1974

Year	All Farms	Farms with less than $2,500 sales
	Percent	
1960	76	339
1961	80	360
1962	88	404
1963	98	463
1964	104	480
1965	108	525
1966	99	570
1967	125	651
1968	129	702
1969	119	762
1970	126	832
1971	146	975
1972	118	993
1973	80	1,298
1974	94	1,584

nonfarm people, not by statistics on the farming operation.

The change from "place" to "establishment" would lead to more information about large scale farming since it would allow the data to be more compatible with the classification used in other economic industries. This change in identifier and the addition of a descripter category of "farms that are operated as a minority part of multi-establishment companies" should improve the ability of the farm data system to reflect the reality of large scale farming.

The addition of more value of sales classes was also to allow the data for the larger farms to be tabulated into more detail. And the use of the type of product descripters in the 1972 SIC manual would provide more detailed tabulations than ever before.

The reduction in number of farms from the changes, approximately 18 percent for the U.S., would vary considerably by state. Some states would lose as little as 3-4 percent while others would lose around 40 percent of what are now counted as farms. The size of farm population, defined as people with residence on a farm, would decrease by the same approximate percentage.

The only potential direct impacts on program funding is the distribution among states of Federal funds for agriculture research, extension and rural development. The total amount of funds available for distribution would not decrease nor would any state lose any of its current share of funds. Only future additional appropriations would be distributed differently with those states having relatively large percentages of farms with less than $1,000 of sales receiving less of these increased funds than they would get if the identifier were not changed. The net effect was very small since the untouched base or current funds are large relative to probable increases in money, at least in the next few years.

The Department of Commerce adopted the new definition after several discussions with their Advisory Committee on Agriculture Statistics. This committee consists of members of the major farm organizations and many farm related industries. The Department of Commerce adopted the recommendation for the following reasons: it would more nearly meet the needs of the principal data users as represented on their advisory committee; it would mean substantial increased completeness of coverage in their mail-out mail-back Census of Agriculture program; and it would mean some reduction in cost for the Census of Agriculture.

Joint press releases were made in August of 1975 by both Departments to announce the change. It was after this announcement that the first major and organized opposition to the change appeared. This opposition came from a few members of Congress who represented states that had relatively larger percentages of small farms, from groups that represented a rural fundamentalism or "farming as a way of life" point of view, and from groups concerned about the rural development policies of the Federal government.

This opposition asked for and obtained Congressional hearings held in November of 1975. Their primary stated concerns about the change centered on the people who would no longer be counted in the statistics on farming. They thought these people might be harmed by no longer being eligible for some programs of USDA, that these people were being "written off" by USDA and Census, that there would be a loss of valuable statistics about these people, and that these people would be harmed by any redistribution among states of Federal funds for agriculture research, extension and rural development. They stated that small family farmers were the backbone of America and that they were afraid the change in definition would add to welfare rolls and urban ills. Illustrative of much of the concern was the statement by one witness that "the absence of a particular group of people from the statistics is synonymous with the denial of the existence of that group and its problems." (6)

The Department of Agriculture and Census Bureau also testified at these hearings and stated again the reasons for making the change. These spokesmen denied any loss of eligibility for government programs or any other potential harm for the people who would be excluded from the economic industry statistics on the farming sector.

A few months later the Congress passed and the President signed a bill which included an amendment to prohibit the change in farm definition before June 30, 1976. This prevented the Census Bureau from using the new definition in their preliminary releases of 1974 data. Further Congressional hearings were held in April and June 1976 at which time testimony was received from a larger number of groups with similar arguments.

Throughout the controversy there has been no opposition to changing the identifer from the concept of place to one of establishment. And the introduction of more descripters has been commended by many of those who oppose change in the minimum sales identifier. Some have stated that more descripters was the only proper way to provide more useful statistics and the identifier should be as all-inclusive as possible.

The current status of all this effort is that a final resolution has not been reached. Those who favor the change and those who oppose it have continued to have discussions.

Lessons From The Experience

What are the lessons that have been learned from five years of study and effort to bring a change in statistical identifiers and descripters into being? First and foremost is that making such a change is a difficult task. The entity called a farm has a "skin" and not cutting into the skin has a very high value to some people. Those people have been part of the "who is affected" group as suggested in the title of this session of the ASA meetings or others representing their interest.

Much of the disagreement has centered on who really is affected and in what way. The Department of Agriculture has said the affected group would be the larger farmers who benefit from existing commercial farm price and income programs. But the lack of opposition from that group or their farm organization representatives leaves open the question of whether they feel they are an affected group. The group who feels they will be affected are the small farmers that would be dropped from the industry statistics and it is their representatives who have opposed the change.

Our observation is that it is not enough to assess who will be impacted by the change and how, but that other groups who think they might be affected also need to be considered. More time could have been given to anticipating what opposition would come forward and what their concerns would be. More of these concerns could have been defused ahead of their being voiced. Realization that change would be so difficult could have resulted in more staff work on proposals to make them more acceptable. More articulate presentation of a wider range of facts and figures could have been constructed. This might have included a special sample survey to find out more about the small farm operators being dropped from the statistics and specific plans to improve other sources of data on these people and their needs for public policies.

Our second observation is that those who get into discussions on an issue like this are not knowledgeable about who benefits from different types of public programs, about what data are used to guide various public policy decisions and about what data are available for this purpose.

The implication is that a good job of providing statistical measures that identify and describe what is really happening in the real world has not occurred. Renewal of data systems has been given too little attention and the misconceptions left by out-of-date systems are difficult to overcome. This is a validation of the concerns on obsolete data systems that were discussed by a committee of the Agricultural Economics profession. (1)

In the discussions on this issue many believed that small farming operations meant low income families and the way to help them was to leave them in the statistics on the farming sector and provide them farm price and income programs. Others were unfamiliar with Population Census data on rural farm and rural nonfarm people, thus pleaded for retaining all possible sources of data. Still others were not aware that the rural nonfarm population, which the operators of small farming operations would become part of as they were dropped from the farm population, was already the largest part of the rural population and several times larger than the farm population. Another commonly held misconception is that the farm population is all the people engaged in farming and only includes those people.

A third observation is that it is easier to get agreement on changes in descripters than changes in identifiers. This raises the question of whether more data system renewal can be accomplished through changes in descripters and lessen the importance of the identifiers. Can statistical information be presented in ways that will cause the public to use data on sub-groups more effectively? Ray Hurley, long the head of the Census Bureau's Agricultural Division, stated in 1962 that none of the descripters or classification systems tried since 1940 had eliminated the misuse of "average farm figures" based on a number that included "more than a million 'not really' farms." This is consistent with Dunn's taxonomy which gives identifiers more importance than descripters. But have we really tried to make sub-categories of data more relevant and useful for policy decisions? Is there a need to stress sub-categories of data to policy makers?

Other questions come out of these recent efforts to change the identifier and descripters for farms. One is whether there should be attempts to create statistical descripters that might make value-laden terms more objective over time. One example in agriculture is the term "family farm." The concept has never been carefully defined or statistically measured against a generally accepted definition but is one of the more powerful terms in debates on farming policy. Another question is whether more constant attention to renewal of data systems and more frequent attempts to make needed changes would help to make change easier. Would more frequent change cause change to be viewed as a more responsible part of the mission for those who maintain the data system? Would more frequent change help bring about the development of more objective criteria to use in deciding when the data system needs renewal? Can such criteria be developed? Or would more frequent change cause too many problems in comparison of statistics over time?

We believe the ultimate lesson from this experience is that renewal of data systems, through continual examination of concepts, identifiers, descripters, and data flows, is a very high priority task in our rapidly changing world. We must have renewed and vigorous commitment to development and maintenance of data systems that meet the needs of the public and private decisionmaker. Trying to cut into the "skin" and make substantive changes in data systems that are badly out of date is a difficult task. It is a task which grows no easier with a lack of attention and concern.

BIBLIOGRAPHY

1. American Agricultural Economics Association, Committee on Economic Statistics. "Our Obsolete Data Systems: New Directions and Opportunities." pp. 867-875, American Journal of Agricultural Economics, Vol. 56, No. 5, Dec. 1972.

2. Bonnen, James T. "Improving Information on Agriculture and Rural Life," pp. 753-763, American Journal of Agricultural Economics,

Vol. 57, No. 5, Dec. 1975.

3. Dunn, Edgar S., Jr. Social Information Processing and Statistical Systems - Change and Reform, John Wiley and Sons, Inc., 1974.

4. Office of Management and Budget, Standard Industrial Classification Manual, 1972, Executive Office of the President.

5. Paarlberg, Don, Testimony before the Subcommittee on Family Farms and Rural Development of the Committee on Agriculture and Subcommittee on Census and Population of the Committee on Post Office and Civil Service, House of Representatives, U. S. Congress, April 26, 1976.

6. Rural America, Inc., Testimony before the Subcommittee on Census and Population of the House Committee on Post Office and Civil Service, House of Representatives, U. S. Congress, June 23, 1976.

7. West, Quentin M., Testimony before the Subcommittee on Family Farms and Rural Development of the Committee on Agriculture, House of Representatives, U. S. Congress, Nov. 7, 1975.

STATISTICAL INFERENCE IN CIVIL RIGHT DECISIONS

Joan G. Haworth, Florida State University and George Mason University
Charles T. Haworth, Florida State University

The Civil Rights legislation of the sixties made illegal a wide variety of actions which had not previously been illegal. The thrust of this legislation centered around the factor of discrimination based upon race, religion, national origin, sex or age. But discrimination is not usually something which can be measured directly; it must be inferred from actions that have been taken. Except for cases where the intent to discriminate and the actions that followed were obvious the courts have, to certain extent, relied on statistical inference to determine the presence or absence of discrimination. In some cases, such as the racial composition of juries, the courts' applications have been reasonable. In other areas, particularly in employment discrimination cases, there have been questionable applications of statistical inference.

Statistics have been used in a wide variety of Civil Right cases including voting, jury selection, congressional reapportionment, housing, testing, worker representation, and wages. Due to the scope of this subject we have not attemted an exhaustive survey. Instead we will focus on the application of statistics in two types of cases: jury selection and fair employment practices. Our emphasis is on the technique used, the underlying model, and the implicit assumptions of the decision.

Introduction

The fact that discrimination cannot be measured directly means that a series of events must be observed. No one act can usually prove discrimination. Just because a member of a minority group is not selected for jury duty, is denied a loan, or is denied a promotion does not indicate the presence of discrimination. It is a pattern of discrimination which the courts look for and a large number of samples showing similar behavior can indicate this pattern. For example, suppose one had a county in which the eligible jury panel was composed of 90% whites and 10% blacks. The fact that a jury, supposedly drawn at random from the panel, was all white would not be unusual--in fact it would occur 28% of the time. However, if the next ten juries were similarly all white, a .0003% probability, it would be reasonable to infer that a nonrandom procedure was involved and that a prima facie case of discrimination in jury selection had been established.

The importance and significance of this type of inference has been recognized. In an oft quoted statement, Chief Judge Brown of the Fifth Circuit Court of appeals noted, "In the problem of racial discrimination, statistics often tell much, and courts listen."[1]

But the fact that the courts have noted the importance of statistics does not mean they always use them in the best manner. Social scientists have long been interested in applying their tools to problems of law but the courts have not been enthusiastic recipients.[2] This lack of acceptance can be attributed to legal training, tradition, and philosophy.[3] Conflicts over facts are generally resolved by hearing both sides and judging the persuasiveness of their testimony. Modern society, however, is bringing more problems involving both complex factual situations and probablistic interpretations into litigation. Many cases involving Civil Rights practices fall into this category.[4] Hence, the use of statistical applications may increase. Let us examine the ways in which statistics have been applied and how better techniques could be used to resolve complex issues in discrimination.

Jury Selection

We start by reviewing the jury selection process. Here the problems are relatively clear cut and the statistical techniques that could be applied are rather simple. One of the earliest issues involving discrimination occurred in the jury selection area. Early decisions established the rule that a conviction in a state court violated the equal protection clause of the fourteenth amendment if it were based on an indictment of a grand jury or a verdict of a petite jury from which blacks were excluded because of their race. The rule was developed in a set of cases in which proof of discrimination wasn't required. In these cases blacks were either excluded by statute or the fact that blacks were excluded from the jury was not contested.

The first case in which probability played a role was the 1934 Supreme Court decision in Norris v. Alabama.[5] Although blacks comprised 7-1/2% of the total male population of the county in which the indictment was brought, no blacks had served on any jury as long as anyone could remember. The trial venire was also challenged because the county where the trial was held was approximately 18 percent black but there also no black had served on a jury within anyone's memory. The Supreme Court was able to infer from these cases that exclusion of blacks was a policy and that discrimination was present. The facts made this decision easy. Similar factual situations existed in other cases in which the courts granted relief, e.g. Arnold v. North Carolina[6] where only one black had served on a grand jury in 24 years. In such cases the court's intuition as to the probability of exclusion of blacks from juries by chance was sufficient basis for judgment.

A more difficult problem was addressed when the issue rested on the underrepresentation but not the exclusion of blacks. Dallas County, Texas was the location of the initial cases in this area. After a murder conviction of a man named Hill[7] had been reversed because of the exclusion of blacks from the grand jury a change in the jury selection process occurred and the next three grand juries each contained one black. A black, Robert Akins, was indicted for rape by the third of these grand juries. Akins appealed on the basis that blacks were underrepresented.[8] The issue rested on the questions as to whether the difference between the expected number of blacks per jury of 1.8552 (approximately 15% of Dallas County's adult male population was black) and the actual number of 1 was a sufficient basis for a finding of discrimination. The court held "we cannot say that the omission...of all but one of the members of a race which composed some fifteen percent of the population alone proved racial discrimination."[9]

The next of the Dallas County underrepresentation cases came 5-1/2 years later. By this time 21 grand juries had been enpaneled since the change in procedures following the Hill decision. Seventeen of these grand juries had one black each while four had no blacks. In Cassell v. Texas[10] a black convicted of murder appealed on the basis of discrimination in the grand jury selection process. The issue apparently considered by the court did not concern the distribution of jurors (1 per grand jury), but the fact that only 6.7 percent of the jurors were black when 15 percent of the adult population was black. However, the state challenged the use of the adult population as being appropriate for the statistical population. They argued that the state law required payment of a poll tax to qualify for jury duty and only 6.5% of the poll tax payee's were black. The court held that no prima facie case of discrimination had been established on the basis of the representation.

A sufficient number of cases have been discussed to give the flavor of the data and issues being faced. Since the descriptions tend to be verbal and based on judicial interpretation, a question might be asked concerning the consistency as well as the accuracy of judicial opinions. A social scientist, Ulmer, set up a theoretical decision making rule for the courts and then tested decisions to see how well they fit his rule.[11] The rule for exclusion cases was that discrimination was found whenever the proportion of blacks was substantial (greater than 7.2%) and they had been excluded over a long period. The rule tested for representation decisions was that the allegation of underrepresentation of blacks would be rejected if their population representation was equal to the actual jury representation at the 95% confidence level. Twenty-five different justices cast a total of 113 votes on jury selection; all but ten were consistent with the listed criteria. Such rulings did not show any liberal-conservative dichotomy. One might question the appropriate level to reject the null hypothesis but the consistency of these applications makes the rule seem to be representative of the decision criteria. Particularly interesting in these decisions was the judicial concern over the proper population.

The decisions discussed above were decided on the basis of judicial intuition as to probabilities. With no representation or in cases of gross injustice the intuition seemed to perform effectively. The problem came when there was some representation. It was clear that there could be some variation from proportional representation. The problem came in distinguishing a significant variation. Here the intuition failed to yield the kind of results that could be obtained with such simple tools as chi square tests. In his excellent discussion of the relevant cases Finkelstein[12] gives numerous cases where proper techniques could have been used to good advantage to judge the random nature of certain factual situations. Let us give one example that illustrates the nature of the cases. He notes a case where the percentage of blacks on grand juries didn't exceed 15% for over 15 years even though blacks comprised 26% of the population. The probability of this happening by chance was 4.63×10^{-21}. Yet the courts did not find discrimination in this case.

Evaluation of Jury Selection Cases

The jury selection cases point out an important feature of applied statistical inference. A model, albeit an implicit one, is necessary to assist the interpretation of data. The model underlying jury selection started with the assumption that there were a set of qualifications, e.g., over 21, and paid poll tax, which defined a qualified pool of individuals. The results of a random draw from this qualified population was compared with the expected value and a hypothesis about the role of discrimination was developed. The model implicitly assumed (quite correctly) that all voters who met certain criteria were equal. Without going into detail we can note that a similar implicit model underlay voter and housing decisions where all individuals who met certain criteria were identical in the eyes of the court.

The courts' intuition about probabilities yielded reasonable judgments in the gross cases. The problem came when more sophisticated methods were needed to test the significance of differences between actual and expected values. Here intuition tends to break down and as a result some rather unlikely events were not held to be prima facie cases of discrimination. However, except for these technical questions the basic method was correct with the implicit underlying model an appropriate vehicle for analysis and a judicial awareness of the presence of chance in any random event.

Fair Employment Cases

Cases under the Title VII Fair Employment Act of 1964 are typically divided into several different areas: seniority determination, promotion procedures, testing, hiring practices, and wage differentials. Upon reflection it is clear that discrimination, if present, should show up in two summary statistics: representation and wages. The other questions such as testing have relevance only when they affect either representation by setting up barriers to hiring or wages by hampering promotions or the job progression tracks.

Let us emphasize the issue in this type of case. If the complaint of discrimination is upheld the employer is liable for either back wages, changes in his normal business practices that will presumably raise costs, or, occasionally, punitive damages. There is a great need here to distinguish between the obligation of the private employer and the obligation of society. Without question society has discriminated against certain minority groups. Whether individual employers discriminated needs to be settled case by case.

Representation

The usual procedure for the plaintiff in a discrimination case in which representation is an issue is to show that there is a difference in the racial composition of the work force, or job classification, from that of the local population. This application is essentially the same as the jury cases. On a few occasions the plaintiffs have been able to rest their cases at this point and the courts have held this racial disparity to be conclusive evidence of discrimination in hiring. Generally, however, a disparity has not been sufficient by itself. It has been enough, however, to establish a prima facie case which the defendant must attempt to rebut. If the defendant is not able to prove a non racial reason for the disparity, discrimination will be found. For example, "the statistics indicate that race is the only identifiable factor explaining the disparity between the jobs held by white employees and those held by black employees."[13]

Examples of such an approach are found in a challenge of the Oakland Police Department where blacks comprised 32-45% of the city's population, but only 3 to 4% of the police force.[14] In this case the court observed that "while such a showing of a significant statistical discrepancy is not in itself dispositive, it is at least some indication that discriminatory forces, albeit subtle ones, may be afoot."[15]

Attempting to prove discrimination by use of a demographic disparity between the general population and the work force assumes that all individuals are equally qualified for a position. We call this the "warm body" hypothesis--and here is where the past effects of discrimination by society need to be carefully separated from the discrimination by an employer. The past effects of discrimination may have reduced the level of qualifications held by members of the minority groups. Such a disparity in qualifications must be corrected, but Title VII should not be used to assign this cost to a private employer. Some courts have recognized this problem while others have not.

A court that was aware of the problem noted that "It is one thing to presume or assume, <u>prima facie</u>--wise or otherwise, that a significant number of a group have the qualifications for schooling or voting, or jury service. It is another thing to assume, prima facie--wise or otherwise, that because a certain number of people exist, be they white or Negro, that any significant number of them are lawyers or doctors, or merchants, or chiefs--or to be concrete, are competent plumbers or electricians, or carpenters."[16]

To a certain extent the courts have understood the fact that there exists a range of qualifications and at least have given lip service to requiring minimum qualifications, e.g., United States v. Household Finance Corp.[17] and United States v. Vepco[18] where the decision gave preferential treatment to minority group members <u>who were qualified</u>.

However, it is clear that if a difference exists in the distribution of qualifications for two groups, proportional representation between the two groups is not an indicator of a lack of discrimination. In fact, proportional representation in such a case would imply discrimination against the members of the better qualified group. A hypothetical example can illustrate this problem. Suppose an employer has 40% black employees in an area where the population is approximately 40% black. Under present interpretations he would appear to be immune from a charge of discrimination on representation since all warm bodies are considered equal. But suppose that the applicants for positions in this firm contain, on average, better qualified blacks. By taking proportional numbers from both races we would end up with the situation where the average black employee was better qualified than the average white. Present interpretation would not consider such a case discrimination.

A case can be cited which illustrates this point. Prior to 1947 blacks were not allowed in major league baseball. After the color barrier was broken teams began to add black players. (The Boston Red Sox was the last team to integrate in 1959.) By 1954 blacks made up 10 to 15 percent of major league players--approximately the percentage of blacks in the general population. Since the distribution of skills varied by race (under proportional representation) the average black player was more highly skilled than the average white player (e.g., blacks had a 40 point higher batting average).[19]

Representation according to proportion without adjustments for qualification (or reference to the appropriate population) can yield incorrect conclusions.

<u>Wages</u>

The same failure to recognize that a distribution of qualifications exists severely flaws some wage cases. The application of a model with an implicit assumption of the "warm body" hypothesis to the analysis of wage discrimination seriously distorts the inferences that can be derived from statistical data.

Typically an individual's contribution to an organization is multifaceted. At one time many factory jobs involved little or no discretion. The "warm body" assumption, although never totally true, might not have been a bad approximation for these factory jobs. However, modern technology has automated most jobs that require no human discretion. Typically work in a modern establishment involves joint production efforts which make certain characteristics desirable to an employer. Interaction with fellow employees and supervisors involves verbal cognitive skills while the complexity of operations often requires the ability to understand written instructions (if for no other reason, than understanding government regulations about the job). A changing work environment makes general knowledge worthwhile. A worker brings different levels of these skills, knowledge and other characteristics to his job.

Economists have formalized the relation between some of the various attributes which a worker brings to his job and his earnings in an "earnings function". An earnings function describes the relationship between a set of attributes, (e.g., age, education, experience, hours worked, etc.) and earnings.

A large number of articles have used these functions to estimate the impact of various factors such as age and education upon earnings. Using a data base made up of individual characteristics and earnings, a function is specified with earnings as the dependent variable. The coefficient on the appropriate independent variables can then be interpreted as the "pay off" for a year of education, or experience, or an extra hour worked, etc. If the coefficients are significant it is presumed that the attribute is desirable or else employers wouldn't pay for the characteristic. Such models have been used to estimate the impact of employee discrimination by including a dummy variable for race. Since the earning differences occurring from differences in worker characteristics are already adjusted for in the equation, the magnitude of the racial variable's coefficient indicates the size of the difference in earnings which cannot be attributed to worker characteristics. Some refer to this magnitude as the amount of discrimination but this isn't quite true since this coefficient is essentially a residual measuring all earnings differences between races not attributed to the variables in the equation. If some other worker characteristic which is correlated with race, like health, were present the amount properly attributed to employer discrimination would be lessened.

Given in Table 1 is an example of an earnings function derived from national data. The results of Table 1 support the propositions which have been developed over time in economic theory. Certain attributes are important to employers. For example, each additional year of education that was completed was associated with an annual increment of approximately $500 in 1969. Similarly age, which is a proxy for experience, is assoicated with higher income until retirement age was reached, etc. Some of the variables were simply corrections for measurement errors induced by cost of living variations across the United States. What is important here is not this particular earnings function but the type of attributes which a market economy rewards with higher compensation.

The contrast between the well established relationship among earnings and worker characteristics found in empirical studies and the relationship used in some cases is disconcerting.

The critical case in this area is Griggs v. Duke Power Co.[20] In his discussion of the case, Chief Justice Burger seemed to take note of the variations.

> Congress did not intend by Title VII; however to guarantee a job to every person regardless of qualifications.
>
> Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such a qualification the controlling factor, so that race, religion, nationality, and sex become irrelevant.

The court then proceeded to do just the opposite of what would have been expected to follow logically. In Griggs the use of educational standards for employment or for promotion were prohibited unless they can be shown to have demonstrable relationship to successful job performance.

At first glance such a standard might appear reasonable. But the unimaginative and literal interpretation which has been associated with demonstrating such a relationship makes it almost impossible. Few employers will be willing to run the risk of meeting the unreasonable standard that has been placed on the employer. For example, in finding a high school diploma requirement discriminatory the court declared "many high school courses needed for a diploma (history, literature, physical education, etc.) are not necessary [for a particular position]".[21] With such reasoning few, if any, degrees could be demonstrated necessary. The inference is that an individual with a 6th grade education is equal to a high school graduate in value to an employer. Such a narrow view implies that literally millions of employers in this country and other countries have been irrational in rewarding educational achievement.

One hopes that the Grigg decision was influenced by the fact that Duke Power had used education as a blatent tool for discrimination. Other decisions offer some hope that years of social science research will not be totally ignored.[22]

Statistical Analysis

Cases involving wage differences could certainly be improved through the use of statistical analysis. Presently the plaintiff presents statistical information on the average wage level of black versus white employees. If the blacks are more senior than whites this data is also included in a three way table. This type of evidence is based on the implicit model that earnings are paid on the basis of factors which are distributed alike in all groups. When seniority is included the model suggests that earnings characteristics are distributed in the same way in the two groups except for seniority. Hence, with this underlying model, a difference by race for a simple average of earnings or an earnings average for each seniority level is all that is required to show that under this standard employers are discriminating. Such a presentation is enough to shift the burden of proof to the defendant who must attempt to explain why such an earnings difference exists.

Usually the defendant's proof focuses on a person by person explanation of the earnings differences. If a more complete model of wage determination had been exhibited early in this case, however, the focus of evidence on both sides could have presented a more cohesive and convincing case.

A complete model of wage determination should be able to determine whether any significant racial differences exist among people with the same education, previous work experience, veteran experience, seniority, absenteeism and other factors which affect people's earnings. This model could be used to test the assumption made by the implicit model that the distributions of all of these characteristics are the same among racial groups. Furthermore, the statistical techniques are available to test this model whenever data exists on the important earnings factors.

Several statistical techniques are available for this type of inference. A set of cross tabulations showing the average earnings by race for all people at a specified seniority level, of

TABLE 1--THE AGGREGATED EARNINGS FUNCTION AND THE SEPARATED EARNINGS FUNCTIONS OF WHITE AND NONWHITE MALES FOR 1969

(t-ratios in parentheses)

Independent Variables	Aggregated Function 1969	Dependent Variable-Individual Earnings Whites Only 1969	Nonwhites Only 1969
Nonwhite	-1427	----	----
	(16.3)		
Age (16-24)	-3732	-3916	-1688
	(38.4)	(37.2)	(8.1)
Age (25-35)	-1753	-1856	- 787
	(23.0)	(22.6)	(4.7)
Age (45-54)	278	380	87
	(3.6)	(3.4)	(0.5)
Age (55-64)	- 2	8	- 469
	(0.0)	(0.1)	(2.3)
Age (65+)	-1503	-1566	-1434
	(9.8)	(9.5)	(4.3)
Schooling	518	546	269
	(66.0)	(63.9)	(16.3)
Rural	-1155	-1198	- 616
	(19.5)	(19.0)	(3.8)
South	- 792	- 753	-1590
	(14.1)	(12.4)	(13.1)
Hours Worked	1.80	1.86	1.34
	(41.7)	(39.2)	(13.7)
Married (SP)	1844	1899	1449
	(22.1)	(20.8)	(8.6)
Once Married	711	733	649
	(5.7)	(5.2)	(2.9)
Nonveteran	- 158	- 169	- 83
	(2.8)	(2.8)	(0.7)
Y-Intercept	-2029	-2448	429
R^2	.32	.31	.29

Source: U.S. Census of Population, Public Use Sample 1970--excerpted from Haworth, Gwartney and Haworth, "Earnings, Productivity, and Changes in Employment Discrimination During the 1960s" The American Economics Review, Vol. 45, March, 1975.

a certain age group, previous experience level, etc. could be used and the resulting differences in earnings by race tested for significance. Alternatively, factor analysis might be used to isolate those variable combinations which seem to measure the principal components of a person's earnings and then could be tested for significant differences by race. Analysis of variance between and among cells which are separated into the relevant categories for each variable could be used to measure significant racial differences in earnings or in the distribution of the earnings' characteristics.

Regression analysis also exists as a technique to show whether there exists significant differences in earnings after we hold the other earnings' affecting characteristics constant. A recent article has suggested that the regression technique is appropriate for determining the presence or absence of discrimination.[23] In several cases with which we are familiar such a technique has been used.[24] In these cases, the same characteristics which were significant in earnings models using national data were significant in the individual firm and for the subset of each race.

The use of regression equations is helpful not only in the establishment (or disestablishment) of discrimination but in the awarding of back pay. After an earnings function for whites has been established the characteristics of each black employee could be substituted in the equation to get an "expected earnings" for each person. Those with a difference between the expected and their actual earnings could be compensated accordingly.

A recent case, White v. Carolina Paper Board Company, illustrates how a more complete model would help the most deserving affected workers. In this case, the employer was found to have discriminated and back pay was to be awarded to compensate the workers for lost earnings caused by the discrimination. All of the workers were awarded the difference in salary between the average white worker and the average black worker.

This back pay allocation scheme did not recognize the fact that some of the black workers had many more years of experience than others. It also ignored the fact that the differences in average earnings by race might be due in part to the presence of skilled craftsmen in the white employee groups. Clearly an earnings function would have been useful in identifying the workers who were most harmed and paying them according to their real ability to earn.

Conclusion

The use of statistics is helpful in many forms of litigation. As courts become involved in more complex phenomena the use of more inferential statistical techniques could be useful. As in most analysis the underlying model and assumptions are important in making the appropriate inference from a body of data. An appropriate model for a jury case might be different from an employment case. Some protocols concerning data and models would be helpful in getting to the essence of the case.[25]

It is clear, however, that the real essence of the statistical analysis question is to identify those factors which influence other factors and measure their significance. In Civil Rights cases, where discrimination cannot be measured or observed directly, statistical analysis is especially helpful in arriving at a fair and equitable decision.

Footnotes

[1]Alabama v. United States, 304 F. 2d 583, 586 (5th Cir 1962).

[2]For a discussion of the role of social science in court decisions on school integration see the Winter 1975 issue of Law and Contemporary Problems which is devoted to the general topic "The Courts, Social Science, and School Desegregation." Of particular interest are "School Desegregation Litigation in the Seventies and the Use of Social Science Evidence: An Annotated Guide." B. Levin and P. Moise, pp. 51-133, "Random Remarks on the Role of Social Sciences in the Judicial Decision-making Process in School Desegregation Cases." J. M. Wisdom, pp. 134-149, and "The Impact of Social Science Evidence on the Judge: A Personal Comment" J. B. Craven, Jr., pp. 150-156 and "Social Science and the District Court" J. B. McMillan, pp. 157-163. Among the judges present there was a lack of unanimity about the role of social science research.

[3]For an insight into the difficulties some lawyers have in accepting a probablistic approach see the reasonable procedure outlined in Finkelstein and Fairley, "A Bayesian Approach to Identification Guidance" 83 Harvard Law Review 439 (1970) and then note the comment by Tribe 84 Harvard Law Review 1329 (1971) which appears to miss the point of the original article. The rejoinder by Finkelstein and Fairley 84 Harvard Law Review 1801 (1971) helps to clarify the distinctions.

[4]So do many of the pollution cases. Courts must evaluate the danger of asbestos in drinking water, for example, on statistical evidence and testimony of a most indirect nature, e.g. the effect of asbestos in the air.

[5]294 U.S. 587 (1935).

[6]376 U.S. 773 (1964).

[7]Hill v. Texas, 316 U.S. 400 (1942).

[8]Akins v. Texas, 325 U.S. 398 (1945).

[9]Id. at 7405-6.

[10]339 U.S. 282 (1950).

[11]Ulmer "Supreme Court Behavior in Racial Exclusion Cases: 1935-1960" 56 American Political Science Review 325 (1962).

[12]Finklestein, "The Application of Statistical Decision Theory to the Jury Discrimination Cases," 80 Harvard Law Review 338 (1966).

[13]Brown v. Gaston County Dyeing Mach. Co. 457 F 2d 1377, 1383 (4th Cir.), cert. denied 409 U.S. 092 (1972).

[14]Penn v. Stumpf 308 F. Supp 1238 (N.D. Cal 1970).

[15]Id. at 1243 n. 7.

[16]Dobbins v. IBEW Local 212, 292 F. Supp 413, 445 (S.D. Ohio 1968).

[17]4 CCH Empl Prac Dec 7680, at 5671 (N.D. Ill 1972).

[18]327 F. Supp 1034, 1042, (E.D. Va. 1971).

[19]For a discussion of this case see Gwartney and Haworth "Employer Costs and Discrimination: The Case of Baseball," 82 Journal of Political Economy 873 (1974).

[20]401 U.S. 424 (1971).

[21]United States v. Georgia Power 474 F. 3d 906, 978 5th Cir. (1973).

[22]The Fourth Circuit has suggested the education, experience, ability, length of service, reliability, and aptitude are proper objective criteria. Brown v. Gaston County Dyeing Mach. Co. 457 F. 2d 1377, 1383 (4th Cir.), cert. denied, 409 U.S. 982 (1972). The Fifth Circuit upheld an employers promotion standard which used prior experience and educational training of the applicants. United States v. Jacksonville Terminal Co. 451 F. 2d 418 (5th Circuit 1971) cert. denied. 406 U.S. 906 (1972).

[23]"Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal," 89 Harvard Law Review 387, (Dec. 1976).

[24]e.g., Ford et al., v. United States Steel Corporation, et al., 371 F. Supp. 1045.

[25]See for example the four protocols suggested in Finkelstein's "Regression Models in Administrative Proceedings," 86 Harvard Law Review, pp. 1442-1475 (June 1973).

A STATISTICAL ANALYSIS OF THE BAIL DECISION

Kathleen B. Brosi and Sidney H. Brounstein, Institute for Law and Social Research

This project has been supported by Grant No. 76-NI-99-0118 awarded by the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

ABSTRACT

The judge's decision to grant a defendant release pending trial is usually made based on two criteria: the likelihood of appearing in court and the likelihood of committing new crimes. Statistical evidence may help the judge make better predictions.

This paper presents descriptive statistics on the types of bail granted, failures to appear, rearrests on bail, and comparative case outcomes for those released on bail versus those not released. It reports the preliminary findings of an analysis of the determinants of the pretrial release decisions, ability to post money bond, and failure to appear.

Under the terms of the D.C. bail law, release on bail is to be based solely on the defendant's likelihood of appearance in court. A separate provision in the bail statute, called "preventive detention," permits denial of pretrial release to certain classes of defendants, based on their dangerousness to the community and strength of evidence in their cases. Designed to protect the community from new crimes rather than to assure court appearances, this provision rarely has been invoked. In requesting high money bond in cases that might otherwise qualify for preventive detention, prosecutors have argued that such defendants are likely to flee rather than face a severe sentence. This paper examines whether judges appear to have been accepting the argument and reports on models being developed to test the validity of that argument.

INTRODUCTION

The increased media reporting in Washington, D.C. of rearrests of defendants previously released pending trial for other offenses has caused a public outcry for reform of the bail system.[1] Of all felony street arrests made in the District of Columbia in 1974, 12 percent involved defendants on pretrial release for prior, unrelated offenses. Under the existing law in the District of Columbia, bail is to be used only to secure the defendant's appearance at court proceedings. In order to protect the community against certain crimes, the District of Columbia has a preventive detention statute [2] which can be invoked in certain circumstances to detain dangerous defendants prior to trial regardless of their risk of flight.

While most defendants in the District of Columbia Superior Court are released on personal recognizance, some are required to post money bond as a guarantee against flight in order to be released. (Money bond is never to be used to prevent danger to the community.[3]) Critics have objected to the use of money bond as unfair to the poor and ineffective to prevent flight.[4]

Pretrial detention of defendants makes it more difficult for them to prepare their cases for trial, strains family relationships, and interferes with their employment. Studies have indicated that even controlling for factors such as prior criminal record, a positive correlation exists between the length of prison sentences and detention prior to trial.[5] In addition to the obvious expense to the defendant, pretrial detention is expensive to the public. In 1962, pretrial detainees in the District of Columbia cost the public $500,000.

The arguments favoring pretrial release are persuasive: cost avoidance, overcrowded jails, fairness, and presumption of innocence. In some cases however, detention is appropriate to prevent flight or harm to the community. The difficult problem is predicting which defendants pose the greatest risks.

Background

In 1963, the Manhattan Bail Project[6] began the development of improved fact gathering for the bail decision and the expansion of release on recognizance. That project devised a set of criteria for release on recognizance. The Manhattan Bail Project led to the National Conference on Bail and Criminal Justice, sponsored by the Department of Justice and the Vera Foundation in 1964. Following this conference, the federal government adopted the Bail Reform Act of 1966.[7]

Under the terms of the Bail Reform Act, the release decision is to be made to insure appearance. Judges are instructed to use the first one of the following conditions which they believe will secure that appearance: 1) personal recognizance, 2) third party custody, 3) restrictions on travel, association or abode, 4) appearance bond (cash) 5) surety bond, 6) other conditions (e.g., night jail). Money conditions are not to be used to assure the safety of any other person in the community.

According to the Bail Reform Act, to determine the conditions of release, the judge is to consider:

1) the nature and circumstances of this offense

2) the weight of the evidence

3) family ties

4) employment

5) financial resources

6) the defendant's character and mental condition

7) past conduct of the defendant

8) length of residence of defendant in the community

9) the defendant's conviction record

10) the defendant's appearance record.

Pretrial (preventive) detention provisions also are included in the D.C. bail law as a means of protecting the community against new crimes that might be otherwise committed by defendants released on bail. Preventive detention is legally available in cases involving dangerous crimes, crimes of violence, and obstruction of justice.

In the first five years under the law, preventive detention has rarely been invoked by prosecutors. The procedure to be followed for preventive detention has been cited as one of the reasons for its infrequent use in practice. First, a motion for preventive detention is made, generally by the United States Attorney's Office. Next, at a hearing held within three days of arrest, it must be shown by "clear and convincing evidence" that this person is eligible for pretrial detention and that there is a "substantial probability" that he committed the offense. The rules of evidence are relaxed at this hearing (e.g., hearsay is admissable). According to the D.C. Bail Reform Act, if pretrial detention is ordered, the trial must be held within 60 days (with few exceptions).

Among the procedural problems cited for not using it more often is reluctance to have the prosecution witnesses testify at the early hearing (possibly subjecting them to intimidation), and excessive case loads, making it difficult to bring the case to trial within 60 days.

Unfortunately, little recent empirical work has been done to help the judges who set bail. The Bail Reform Act includes a list of factors that a judge may consider in setting bail. It does not provide specific guidelines on how those factors are to be used. Undoubtedly, through experience, judges develop their own schemes for weighing these factors. Since each judge independently develops a method for dealing with the factors, based on the cases he sees, it is likely that different judges set bail differently.

The bail decision results in the incarceration of more defendants than does the sentencing decision. Hence, any disparities that might exist in bail release conditions would have even more negative impact on evenhandedness in the criminal justice system than disparities that might exist in sentencing.

Purpose of Research

The purpose of the research described in this paper is to provide policy and decision makers with empirical information on the functioning of the bail system. Judges tend to be reluctant to use statistics, since each case must be decided ultimately on its individual merits. The thesis of this research is that there is no conflict with that philosophy. The judge at the bail hearing is faced with making a prediction of the defendant's future behavior (failure to appear in court or rearrest while on conditional release). He has to consider the conditions of the individual case before him. But if in addition, he is provided with statistical information to help him make a better prediction, the interests of justice and the community will be better served.

Data Base

The data to be used for the empirical analysis of pretrial release have been collected mainly by PROMIS (the Prosecutor's Management Information System) which records over 170 items of data for each case. PROMIS has been operating in the Superior Court Division of the U.S. Attorney's Office of Washington, D.C., since 1971. In addition to characteristics of the defendant, the criminal incident, and the case, PROMIS records court actions, including the release decision of the judge made at the initial court hearing after arrest (arraignment), and the issuance of bench warrants for failure of the defendant to appear. Since PROMIS data are maintained on all U.S. Attorney's cases, information is available on subsequent rearrests of persons released while awaiting trial.

This paper presents descriptive statistics on the types of bail granted, failures to appear, and rearrests on bail. It reports the preliminary findings of an analysis of the determinants of the pretrial release decisions, the ability to post money bond, and failure to appear.

STATISTICAL TECHNIQUES

One way to learn about causal relationships is to conduct a controlled experiment. Because of ethical and legal considerations, however, controlled experiments have very limited applicability in the criminal justice system. But one can gain insights into causal relationships by applying multivariate techniques to nonexperimental data (i.e., data that accumulate in the normal course of operations). Nonexperimental data are becoming more abundant in the criminal justice system, particularly due to the growth of automated data processing systems such as PROMIS.

The primary analytical tool used in this research project is regression analysis. Regression analysis is used not only to predict unknown values of certain variables based on the known values of other variables, but also to describe relationships between variables, so that inferences about causality are possible. It is this latter use of regression analysis that makes it particularly well suited for application within the bail study since it is the effect that given changes in policy factors have upon each of several different output or performance measures that is to be examined. For example, when everything else is held constant, what effect does an increase or decrease in the percentage of defendants released on personal recognizance have upon the failure to appear rate? To what extent, in turn, are release rates affected by the type of defense counsel appointed by the court? Multiple regression analysis enables these types of questions to be addressed, even though all other factors may not, in reality, have been held absolutely constant. Assuming that the direction of causality is at least partly from independent to dependent variable, the analysis, when properly applied, measures the observed effect that each independent variable has upon the dependent variable, after taking account of the effects that all of the other independent variables have upon the dependent variable.

When multicollinearity exists, these subrelationships can also be analyzed using multiple regression. In such cases, the full model will consist of more than one regression equation. For example, in estimating the effect of the amount of money bond set on the likelihood of failure to appear, it may be that the causal relationship between these two variables is likely to run in the opposite direction as well. Indeed, the judge is supposed to set money bond in order to assure the defendant's appearance in court.

There are two basic types of complications in isolating effects through regression analysis: <u>circularity</u> and <u>recursivity</u>. Circular relationships, such as that between the amount of money bond and the likelihood of failure to appear both start and end with each factor involved in the circle. Recursive relationships involve a chain of causally related factors, with one affecting the next in a noncircular manner. For example, consider the factor "employment status" of the defendant. Presumably, the employment status affects the type of bail condition set and the latter affects probability of rearrest while released on bail. The set of equations comprising the models for analyzing failure to appear and probability of rearrest while released on bail will be structured to sort out these effects.

When first developed, regression analysis was applicable only to systems that were linear, or nearly so; also, it was used only to relate unbounded "scalar" (i.e., measurable) variables; further, it was suitable only when relationships were homoscedastic (i.e., the property that exists when the distribution of the dependent variable on each independent variable has a constant variance for all values of each independent variable). Now, due largely to extensions developed within the discipline of econometrics, regression analysis has been made capable of overcoming each of these earlier limitations. Nonlinearity can be dealt with adequately in many cases by transforming to the linear form those variables that are nonlinear with one another (e.g., by regressing the dependent variable on the logarithm or exponent of a nonlinear independent variable). "Taxonomic" (i.e., qualitative) or other discrete-valued variables, and combinations thereof, can be used as independent variables in a regression equation through the creation of "dummy" (i.e., binary or "zero-one") variables.

A dichotomous dependent variable can be converted to a variable suitable for regression by aggregating individual observations into cells, expressing the dependent variable as a proportion, transforming the proportion to a suitable form (e.g., logit and probit transformations), and regressing this transformed variable on the independent variables. And homoscedasticity can usually be imposed through an "analysis of residuals," followed by the application of an appropriate scheme for weighting the observations.

The refinements of regression analysis noted in the above paragraphs are presented in H. Theil's Princ[illegible]s of Econometrics[8] and A. S. Goldb[illegible]er's Economic Theory.[9]

STATISTICAL MODE[illegible] FOR PREDICTING THE BAIL DE[illegible]SION AND OUTCOMES

Statistical models using factors suggested in the D.C. bail law and other case and demographic factors are being constructed to predict:

1) failure to appear (bench warrants issued)

2) rearrest

3) whether or not a defendant will be able to post money bond to obtain release

4) conviction or acquittal given the bail information.

The variables to be considered as predictors are:

1) past record of the defendant (arrests, prosecutions, and convictions separately)

2) seriousness of the instant (current) offense (Sellin-Wolfgang score,[10] maximum sentence associated with the most serious charge in the case)

Demographic characteristics of defendant

3) age of the defendant

4) sex of the defendant

5) race of the defendant

6) employment status of the defendant

7) length of residence of defendant in the District of Columbia

Case characteristics

8) number of codefendants

9) pending cases

10) type of victim

Case processing characteristics

11) type of defense counsel (Public Defender Service, Criminal Justice Act, Retained, other - attempt to look at intensity of communication)

12) time from release to bench warrant or new offense

13) elapsed time from arrest to final disposition

Indicators of strength of evidence

14) number of lay witnesses

15) whether tangible evidence recovered

16) delay from offense to arrest

17) victim-defendant relationship (stranger, family, etc.)

The dependant variables are being constructed to reflect outcomes whose probabilities may be affected by the bail decision (such as failure to appear), and the preventive detention decision (such as rearrest). These models are designed to assist in determining 1) the minimum conditions required to assure appearance, and 2) the types of defendants likely to be rearrested (and therefore deserving more careful scrutiny when bail is set).

The models to be developed are:

1. Prediction of failure to appear (the issuance of a bench warrant)

One hypothesis is that the Vera criteria,[11] widely accepted as predicting failure to appear (prior convictions, unemployment, lack of community ties, etc.) for pretrial release, will actually select those candidates most likely not to appear in court as scheduled. A second is that the same conditions that are required for preventive detention (strong evidence, serious crime, serious criminal record) predict failure to appear (FTA). In order to focus on the effects of variables under the judge's control, the [illegible] will attempt to isolate the effects on FTA of the pretrial release decision itself, controlling for other variables such as whether the defendant was able to make bond and secure release.

2. Prediction of rearrest of the dangerous defendant on bail

One hypothesis to be tested is that the likelihood of rearrest of dangerous defendants on bail is a function of the defendant's prior arrest record, age, crimes charged, and the likelihood of pretrial release.

3. Prediction of the decision to grant release conditional on money bond

The hypothesis to be tested is that judges are more likely to grant release conditional on money bond to defendants with more serious prior criminal records, who face more serious charges with stronger evidence, independent of the Vera criteria and prior FTA incidents. The implication of such a finding would be that judges may be setting money bond conditions in lieu of invoking preventive detention.

4. Prediction of whether or not the defendant will make money conditions of release

The hypothesis is that defendants with more recidivistic prior criminal records for property crimes are more likely to make money conditions of release, controlling for factors such as the amount of money bond set.

5. Analysis of the effect of pretrial release on case disposition

The hypothesis is that defendants detained prior to trial are more likely to be convicted through a plea or trial, controlling for factors such as prior criminal record and crime seriousness.

DESCRIPTIVE STATISTICS ON BAIL PRACTICES

The first stage of the research was designed to develop descriptive statistics of pretrial release practices in order to obtain preliminary insights regarding the nature of the data. These statistics are based on a total of 17,534 arrests brought to the Superior Court Division of the U.S. Attorney's Office for the District of Columbia. This Division operates as the local (as distinguished from Federal) prosecutor of serious misdemeanors and all felonies charged in the District of Columbia.

Types of Pretrial Release Granted at Initial Court Hearing

In 1974 in the D.C. Superior Court, 58 percent of all defendants were released on their own recognizance at the initial hearing. Another 12 percent were released for supervision by third party custody programs. Financial conditions of release were set in 27 percent of the cases. Of these, 20 percent were released on surety bond (a bail bondsman must guarantee the bond) and 7 percent were released on cash bond (the defendant guarantees the bond). In a sample of robbery and burglary cases, 29 percent of the defendants who had financial conditions set were able to secure release by posting bond.

Another source[12] has reported that of all defendants (misdemeanors and felonies) who had financial conditions imposed, 75 percent eventually made bond. This figure may be consistent with the 29 percent rate for robbers and burglars, since it includes many misdemeanor defendants who tend to have lower bond amounts set and consequently are more likely to obtain release. Also, robbers and burglars tend to have more extensive criminal histories than other types of felons, which would indicate higher amounts of bond set.

Persons accused of misdemeanors were given less stringent conditions of release than were persons accused of felonies. Table 1 presents the various types of release granted at the initial court hearing in 1974. Nearly 80 percent of the misdemeanor defendants were released without financial conditions at the initial hearing, while 61 percent of the felony defendants were so released.

The type of bail granted at the initial hearing also varied by crime type. Table 2 presents release type granted for murder, rape, assault, robbery and burglary cases. Surety bond is the most common condition of release set in murder cases. Personal recognizance is granted in an additional 16 percent. That means that about half the murder defendants are released from custody at their initial hearing. The other half are detained on financial conditions.

Sixty-nine percent of the rape defendants, 75 percent of the aggravated assault defendants, 60 percent of the robbery defendants, and 64 percent of the burglary defendants were released without posting money bond at their initial hearings. The remainder were detained on financial conditions, primarily surety bond.

DETERMINANTS OF THE JUDGES' BAIL DECISIONS

The pretrial release decision is made by the judge in a very hectic atmosphere. One judge sitting in arraignment court on a typical day makes all the release decisions in about fifty cases.

TABLE 1

TYPE OF BAIL GRANTED AT INITIAL HEARING IN 1974 (percentages)

	MISDEMEANORS (N=6,485)	FELONIES (N=5,061)
PERSONAL RECOGNIZANCE	70	44
THIRD PARTY CUSTODY	9	17
CASH BOND	7	8
SURETY BOND	13	29
OTHER (includes mental observation, rehabilitation for alcoholics, and special conditions)	1	2

NOTE: Percentages are based only on cases for which release type was recorded. The data presented here excludes 18 percent of the misdemeanors and 17 percent of the felonies because the type of release was not recorded for them.

TABLE 2

RELEASE TYPE GRANTED AT PRESENTMENT BY CRIME CATEGORY (percentages)

	MURDER (N=129)	RAPE (N=145)	AGGRAVATED ASSAULT (N=865)	ROBBERY (N=1,244)	BURGLARY (N=924)
PERSONAL RECOGNIZANCE	33	39	64	39	46
THIRD PARTY CUSTODY	16	30	11	21	18
CASH BOND	5	2	4	7	8
SURETY BOND	44	23	18	32	25
OTHER (includes mental observation, alcoholic treatment, and special cases)	2	6	3	1	3

NOTE: Percentages are based on cases in which release type granted was known. The data presented here do not include approximately 17 percent of the cases in which the release type granted was not recorded.

He has available to him information provided by the D.C. Bail Agency on the age of the defendant, his residence, his family and employment status, the number of times he has failed to appear at court proceedings, and the number of times he has been convicted. The Bail Agency recommends the type of bail to be granted to the defendant. The judge also hears recommendations from the prosecutor and defense attorney.

A preliminary analysis explored possible determinants of the judges choice between "hard" bond (i.e., cash or surety) and "soft" bond (i.e., personal recognizance or third party custody). This preliminary analysis investigated the relationship between the judge's decision and the following variables:

- sex of the defendant
- age of the defendant
- residence of the defendant (local area or not),
- type of defense counsel (Public Defender Service, other),

Case and case processing characteristics

- seriousness of the offense (as measured by the Sellin-Wolfgang index of crime seriousness),[13]
- prior record of the defendant (as measured by the Base Expectancy Scale, Gottfredson Score of defendant seriousness),[14]
- number of codefendants in the case,
- number of witnesses,
- type of victim (i.e., person or institution),
- whether or not the defendant had another pending case at the time of this arrest.

Other control variables (i.e., employment, prior failure to appear, conviction record) are required and will be included in a more complete analysis to be conducted in the next phase of this study.

An ordering of the variables that were found to explain the decision to impose financial conditions was made.[15] The highest ranking variable of those tested in explaining the imposition of financial conditions was the Gottfredson score (a measure of the defendant's prior criminal record); the more serious the defendant's score, the more likely it was that he would have either cash or surety bond set at his initial hearing. The age of the defendant was the second highest ranking of the variables tested; the older the defendant, the more likely it was that he would have hard bond imposed. Third in rank was the existence of a pending case. If the defendant had another open case (and, therefore, was already on some form of pretrial release) at the time of his arrest in this case, he was more likely to have financial conditions imposed. The seriousness of the offense (as measured by the Sellin-Wolfgang index) was next in order of impact. The more serious the case, the more likely that hard bond would be set. A measure of strength of evidence against the defendant--the number of witnesses identified by the police officer--was also significant. Imposition of financial conditions became more likely as the number of witnesses increased.

Based on this preliminary analysis, it appears that the judge, in deciding between personal recognizance and money bond, is most influenced by the prior criminal record of the defendant, but he is also concerned about the seriousness of the instant offense (the current case under consideration) and the strength of evidence against the defendant. The amount of the money bond set appears to be most influenced by the seriousness of the offense. The more serious the offense (as measured by the Sellin-Wolfgang index) the higher the amount of money bond. The criminal history of the defendant is also an important determinant of the amount of the bond. The amount of money bond set increases with the seriousness of the defendant as measured by the Gottfredson score. Apparently judges who set money bond consider first the seriousness of the current offense, and next the prior record of the defendant. This may indicate that judges are accepting the rationale that a defendant with a more serious prior criminal record, facing a serious charge with strong evidence, is more likely to flee. On the other hand, it may indicate that high money bond, in lieu of invoking preventive detention, is being used to protect the community from dangerous releasees.

SAMPLE STUDY OF RELEASE ON FINANCIAL CONDITIONS

1. Time From Imposition of Financial Conditions to Release

Financial conditions of release are set at the initial hearing in 27 percent of all cases. PROMIS does not collect changes in release conditions that occur after the initial decision; if one were to rely exclusively on PROMIS data, it would not be possible to determine whether defendants held on cash or surety bond were subsequently able to post bond and be released, thus having the opportunity to recidivate or flee.

As part of an analysis of robbery and burglary cases, a 50 percent random sample of those defendants who had financial conditions of release set at the initial hearing was drawn. A manual search of court records for those 464 cases was conducted to collect data on subsequent changes in financial conditions and defendant release status and merge it with PROMIS data on those cases. (Currently, data on a random sample of all cases--including crimes other than robbery and burglary--are being collected to investigate these issues.) The robbery and burglary sample study disclosed that about 29 percent initially detained on financial conditions were eventually released. The median elapsed time from the initial court hearing to release was four days for defendants who were eventually released. Table 3 shows the distribution of elapsed time from the imposition of financial conditions to release for those defendants who are released prior to trial.

2. Which Defendants on Financial Conditions are Released

What factors explain why the 29 percent were able to make bond and the other 71 percent were not? Such an explanation requires a multivariate analysis.

A preliminary analysis of the effect of the following variables on the release from financial conditions was conducted:

- number of prior arrests of the defendant
- residence of the defendant
- type of victim
- number of witnesses identified by police
- seriousness of the defendant criminal record (as measured by the Gottfredson score)
- whether cash or surety bond was set
- number of bond changes
- age of the defendant
- type of defense counsel (PDS, other)
- seriousness of the offense (as measured by the Sellin-Wolfgang index)
- number of codefendants
- whether property or other tangible evidence was recovered
- final bond amount.

If the final condition was a cash bond (rather than a surety bond) the defendant was substantially more likely to be released. This may be because a 10 percent refundable deposit of the bond money with the court is often sufficient for release on cash bond. As expected, the greater the number of times that the bond conditions were changed and the

TABLE 3

TIME FROM IMPOSITION OF FINANCIAL CONDITIONS TO PRETRIAL RELEASE

NO. OF DAYS DETAINED	NO. OF DEFENDANTS RELEASED	PERCENTAGES
0	30	25
1-4	33	27.5
5-8	14	11.7
9-12	9	7.5
Over 12	34	28.3
TOTAL	120	100

lower the dollar amount of the bond, the greater the likelihood that the defendant would be released. And the more serious the defendant's prior criminal history, the less likely he would be able to post bond and be released.

ANALYSIS OF FAILURE TO APPEAR

Several preliminary multivariate analyses are being conducted in an attempt to identify characteristics of defendants, cases, and offenses which are related to the issuance of bench warrants as a measure of failure to appear in court. (A simultaneous equations model is being constructed to sort out the effects of the determinants of the pretrial release decisions on failure to appear.)

One of the challenges in analyzing failure to appear is to develop logical comparison groups that have equal opportunity for failures. Felonies and misdemeanors have different numbers of scheduled court appearances, and the bail decision itself affects the defendant's opportunities. In conducting these preliminary analyses, the cases were initially separated into felonies and misdemeanors; then broken out into cases of those defendants released on personal recognizance; those released on third party custody; and those actually released after meeting financial conditions. The following variables were then tested as determinants of the issuance of bench warrants:

- Whether or not the defendant had a pending case at the time of this arrest
- Whether or not the defendant was on probation or parole at the time of screening of this case
- The number of lay witnesses identified at the time of screening of this case
- The length of time (in days) between the offense and the arrest
- The number of codefendants in the case
- Whether or not property or other tangible evidence was recovered
- Whether or not the victim of the offense was a business or institution
- The Sellin-Wolfgang case seriousness index
- The number of prior arrests of the defendant
- The Gottfredson defendant seriousness score
- Age category of defendant
- Whether or not the defendant was a resident of the local area
- Whether or not the defendant was known to be a drug user
- Whether or not the defendant and the victim were strangers

In analyzing 3072 felony cases involving defendants released on all forms of personal recognizance, with and without conditions, an additional independent variable "whether or not the defendant was released on third party custody" was included. If the victim of the offense was an institution rather than an individual, the case was more likely to result in a failure to appear. Cases involving institutional victims have a higher likelihood of conviction, suggesting that the stronger the evidence, the more likely the defendant will fail to appear.

The more serious the offense, the less likely the defendant was to fail to appear. This might suggest that those defendants facing less serious charges anticipate that the police will not come looking for them to execute the bench warrant. Those cases in which the defendant and victim were strangers were more likely to result in issuance of a bench warrant.

An analysis of a sample of 350 cases involving defendants released on financial conditions after posting bond revealed three variables to have a significant effect on failure to appear among all those tested. The final amount of money bond set had the strongest effect, but it went in the opposite direction to its intended purpose. Among those defendants who succeeded in obtaining release, the higher the bond the defendant had to post, the more likely he was to fail to appear and forfeit it. This puts into question the premise of setting high money bond, unless the intent is to detain the defendant by keeping bond so high that he can't post it. Cases involving male defendants were less likely to result in issuance of a bench warrant, while cases in which the defendant and victim were strangers were more likely to result in issuance of a bench warrant.

It was desirable to test the hypothesis that delay increases the likelihood of failure to appear. As mentioned above, cases vary in the number of court appearances that are scheduled. The large proportion of cases that are terminated before trial due to plea bargaining and dismissals makes it difficult to construct study groups of cases with comparable opportunities for FTA. Also, there is likely to be a circular relationship between failure to appear and delay. The variable "elapsed number of days between arrest and postindictment arraignment" was constructed for the group of indicted felonies. Cases where there was a preindictment failure to appear were excluded from this preliminary analysis to make sure the delay preceded the failure to appear. Based on 1539 indicted felony cases in which defendants were released on all forms of personal recognizance, the variable showing the most significant effect on failure to appear was the delay between arrest and postindictment arraignment. This suggests that court delay does indeed increase the chances of failure to appear in court. Those cases involving male defendants and recovered property or evidence were more likely to result in a failure to appear. Cases with less serious charges were more likely to result in issuance of a bench warrant.

In analyzing indicted felony cases of defendants released on financial conditions after posting bond, the strongest determinant of failure to appear was court delay. The higher the final bond amount the more likely there would be a failure to appear. Cases involving victims and defendants who were strangers were more likely to result in issuance of a bench warrant. Cases involving male defendants and cases involving codefendants were less likely to result in issuance of a bench warrant.

CRIME WHILE RELEASED ON BAIL

In 1974, 12 percent of all felony arrests involved defendants who were on pretrial release from prior separate and distinct criminal arrests at the time of their arrests. Another 14 percent of the felony arrests involved defendants who were on other forms of conditional release (i.e., probation and parole). Nearly one-third of the robberies and burglaries in 1974 involved conditionally released defendants. Table 4 shows the percentage of arrests involving conditionally released defendants for burglary, murder, rape, robbery, and assault. Nearly one-third of the robberies and burglaries in 1974 involved conditionally released defendants.

Another question relates to the seriousness of crimes committed while the defendant is on bail. The largest category of pending cases for defendants arrested for robbery while on bail was robbery. In burglary cases as well, the largest category of pending crime was burglary. But many other categories of crime were included in the pending cases for these robbery and burglary defendants. An analysis of the shift in crime type and crime seriousness for defendants released on bail who are rearrested will be conducted in the next phase of this project.

RELATIONSHIP OF BAIL STATUS TO CASE DISPOSITION

Critics of the bail system maintain that pretrial incarceration forces guilty pleas from defendants who would not otherwise plead guilty, and increases the likelihood of conviction and incarceration. A resolution of that issue must wait for a more sophisticated analysis than this descriptive profile. Bivariate statistics show similar plea rates for those detained versus those not, and a higher conviction rate for detained defendants (see Table 5). These statistics are not conclusive, because it is necessary to control for other factors, such as seriousness of the crime, strength of evidence and prior criminal record. (It may be that judges are setting release conditions based on the strength of evidence against the defendant so that defendants who are more likely to be convicted have more severe conditions of release set initially.)

TABLE 4

CONDITIONAL RELEASE AT TIME OF ARREST (1974)

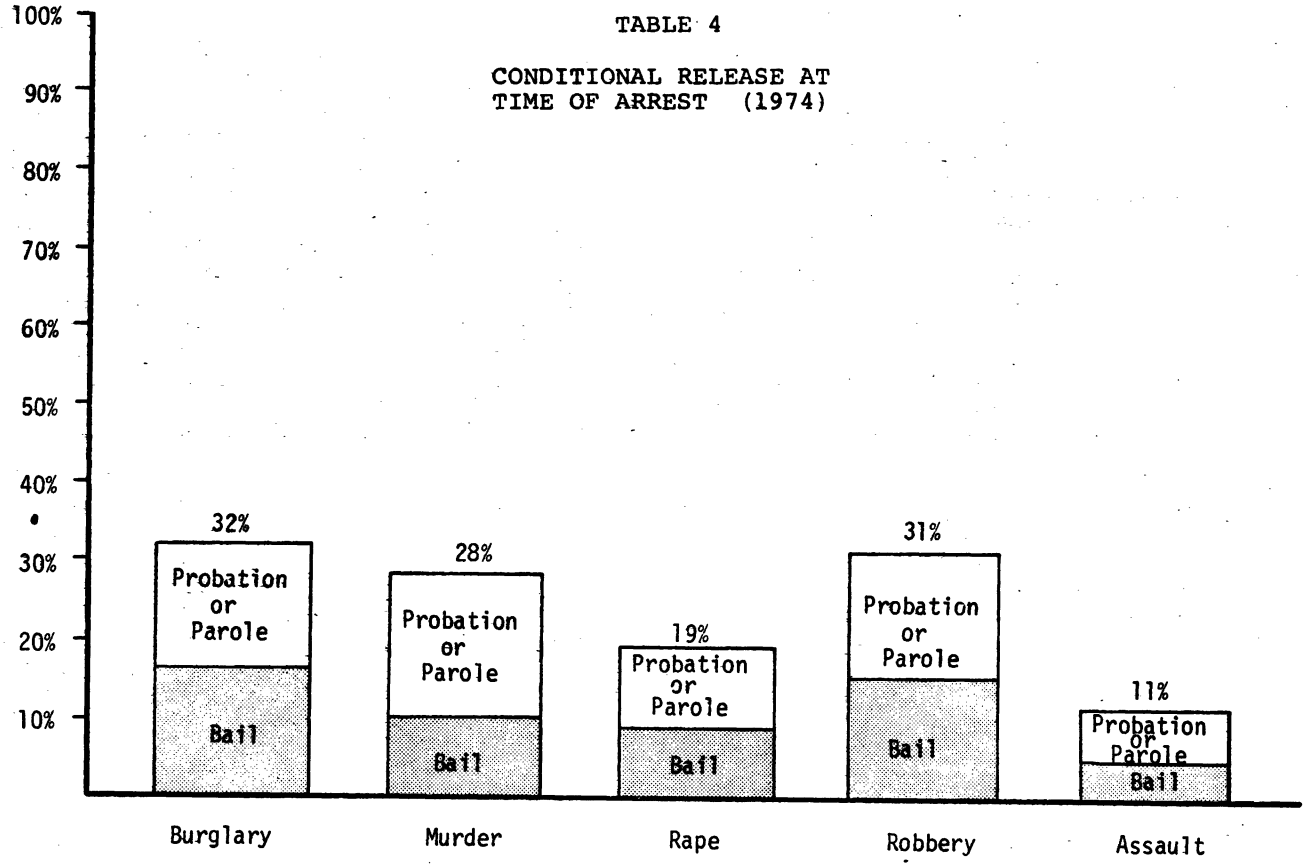


TABLE 5

FINAL DISPOSITION BY WHETHER OR NOT DEFENDANT MADE BOND

	PLED	CONVICTED	ACQUITTED	NOLLED OR DISMISSED	OTHER	OPEN	
DEFENDANT MADE BOND (was released)	34% (45)	7% (10)	6% (8)	40% (54)	2% (3)	10% (14)	N=134
DEFENDANT DID NOT MAKE BOND (was detained)	35% (117)	12% (41)	3% (10)	40% (132)	2% (6)	7% (24)	N=330

CONCLUSIONS

Based on the preliminary analysis of the determinants of the decision to set financial conditions of release rather than release on personal recognizance, it appears that the judge is most influenced by the prior criminal record of the defendant, and to a somewhat lesser extent, by the seriousness of the instant offense and the strength of the evidence. Older defendants and those with other cases pending against them were also more likely to have financial conditions of bail set. When money conditions were set, the amount is influenced most by the seriousness of the offense and next by the criminal history of the defendant. Therefore, there is some indication that judges have been accepting the prosecutor's rationale that defendants who have more serious offenses with strong evidence are more likely to flee. On the other hand, it may indicate that high money bond is being used in lieu of preventive detention to protect the community from dangerous releasees.

Preliminary findings of an analysis of the determinants of failure to appear suggest that cases involving more delay, stronger evidence, and a stranger-to-stranger relationship between the defendant and the victim were more likely to result in the issuance of a bench warrant. No support was found for the hypothesis that defendants with more serious prior criminal records and charged with more serious crimes are more likely to fail to appear.

A sample analysis was made of the likelihood of securing release by posting money bond. Defendants released on cash bond rather than surety, those who had their bond conditions changed most often (suggesting the importance of a good defense counsel), those with lower amounts of bond (not surprisingly), and those with less serious criminal records were more likely to be released.

Although the results reported in this paper are preliminary, they suggest that statistical information can be provided to judges to help them make more informed bail decisions.

Footnotes

[1]Maurice J. Cullinane, "Stopping Career Criminals," Washington Post, April 15, 1976. Timothy S. Robinson, "Repeaters Cause Big Share of Crime," Washington Post, April 10, 197-.

[2]23 D.C. Code, § 1322.

[3]Jones v. U.S., D.C. App. No. 9961, November 5, 1975.

[4]Foote, The Administration of Bail in New York City, 106 U.Pa.L.Rev. 693 (1958).

[5]Rankin, The Effect of Pretrial Detention, 39 NYU.L.Rev. 641 (1964).

[6]Ares, Rankin and Sturz, The Manhattan Bail Project, 38 N.Y.U.L.Rev 67 (1963)

[7]23 D.C. Code, § 1321 et. seq. (1970).

[8]Theil, Principles of Econometrics (New York: John Wiley & Sons, 1971).

[9]Goldberger, Econometric Theory (New York: John Wiley & Sons, 1964).

[10]See PROMIS Briefing Paper No. 3, Uniform Case Evaluation and Rating, Institute for Law and Social Research, Washington, D.C., July 1975.

[11]Ares, Rankin and Sturz, The Manhattan Bail Project, 38 N.Y.U.L.Rev 67 (1963).

[12]Lewin and Associates, An Evaluation of Third Party Custody Programs, Washington, D.C., October 1975.

[13]See PROMIS Briefing Paper, op. cit.

[14]Ibid.

[15]This ordering is based on ranking the significant (from t-tests) variables on their elasticities. The elasticity of one variable, say y, with respect to another, x, is a number indicating the percentage increase (a negative number indicates a decrease) in y that results from a 1 percent increase in x. In other words, elasticity is a measure of the impact or effect of change in one variable on another variable.

A PROBABILITY MODEL OF A PYRAMID SCHEME

Joseph L. Gastwirth, The George Washington University

Periodically, the pyramid or "chain letter" scheme is offered to Americans under the guise of a business dealership. Recently, Glen Turner's Koscot Interplanetary Cosmetics firm has been charged with pyramiding by the FTC, SEC and various state agencies [2]. The total loss to the public has been estimated to be 44 million dollars. The promoters offer people a dealership or sales job in which most of their remuneration comes from recruiting new dealers (or salespersons). The basic fraud underlying a typical pyramid scheme is that every participant cannot recruit enough other people to recoup his investment, much less make a profit, since the pool of potential participants is soon exhausted.

The usual method of prosecuting such schemes is to show that if the representation of the promotional brochures were valid (e.g., members could recruit two new people a month), then within a short period of time (about 18 months) the entire population of the U.S. would have to participate. Thus, the last members would have no one to recruit. Although this argument based on the geometric progression is sometimes rejected by courts as unrealistic [3], pyramid scheme operators have placed a quota (or limit) on the number of participants in a specific geographic area in order to evade this line of prosecution. This article develops a probability model of this quota-pyramid scheme and the following results which also apply to unlimited schemes are derived:

1. The vast majority of participants have less than a 10% chance of recouping their initial investment when a small profit is achieved as soon as three people are recruited.

2. On the average, half of the participants will recruit no one else and lose all their money.

3. On the average, about one-eighth of the participants will recruit three or more people.

4. Less than one percent of the participants can expect to recruit six or more new participants.

While the above results can be "approximately" derived by ordinary limit theorems, for purposes of legal cases an absolute statement that a probability is small is more useful than an "approximate" statement. Thus, the above results are derived from a new probability bound on the sum of "small" binomial r.v.'s which is related to previous work of Hodges and LeCam [4].

Description of One Pyramid Scheme

A recent legal case in Connecticut [6] illustrates the confounding of legitimate business enterprise with a pyramid operation. People were offered dealerships in a "Golden Book of Values" for a fee of $2500. In return for their investment dealers could earn money in two ways: In each geographic area dealers were to develop a "Book of Values" for eventual sale to the public. First, they were to sell advertisements to merchants for $195 apiece and could keep half as a commission. Each advertisement offered a product or service at a discount, so that a "Book of Values" containing 50 to 100 discount offers could be sold to the public. The public was to pay $15 for the Book of Values, of which dealers were to keep $12. Second, a dealer had the right to recruit other dealers and was to receive $900 for each new recruit. Since the creation of a complete "Book of Values" for sale to the public takes a substantial amount of time, clearly the recruitment of new dealers is the most lucrative aspect of the venture.

In the recruitment brochure the possibility of earning large sums of money was illustrated by the following example: A dealer will bring people to weekly "Opportunity Meetings" and should be able to enroll other dealers at the rate of two per month. Thus, at the end of one year, the participant should receive $21,600 from the recruitment aspect alone. The prosecution showed that this misrepresents the earnings potential by asking the following question: "Suppose dealers who are enrolled can enroll two other dealers per month; as time went by, what would happen?" Professor Margolin (of Yale) testified that there would be a tripling of the number of dealers per month and by the end of 18 months, the geometric progression would exhaust the population of the United States. Clearly the cited recruitment brochure is misleading as all participants cannot come close to earning the indicated amount of money.

The "Golden Book of Values" pyramid system had an extra statistical nuance; i.e., there was a quota of 270 dealerships for the State of Connecticut. The Court noted that if each new dealer was successful in recruiting two dealers per month, only 27 would make a profit and the other 243 would lose money depending on how far down the pyramid they were.

Since a real pyramid operation would not be as regular as the Court described it, i.e., even at the beginning every participant would not enroll exactly two new dealers each month, in the next section we develop a probability model of the pyramid scheme. The model enables us to calculate the probability distribution of the number of people each participant will recruit and realize how strongly the probability of recouping one's initial investment depends on <u>when</u> the participant enters the pyramid scheme. Furthermore, the fraction of participants who can expect to recruit no one, can be derived.

<u>Calculating the Expected Return and Probability of Earning a Profit for Individual Participants in a Quota Pyramid System</u>

Economists evaluate the profitability of a business venture by comparing the initial investment to the "expected return" over a period of time. Suppose one is offered the opportunity to pay c dollars to enter a pyramid scheme which will terminate when the total number of participants is N where the fee for finding a new recruit is d dollars. Should one join? The answer is yes only if the expected number of people one will recruit, say R, is greater than c/d, i.e., one's expected earnings (Rd) are larger than the cost (c) of entering the plan. In this section we calculate the expected number of people the k^{th} participant will recruit assuming that all current participants have the same chance of recruiting the next member.

For ease in exposition we focus on the k^{th} entrant into the system. Since there are now k participants, each of whom presumably is recruiting, the probability that any particular one of the k current members recruits the next one is 1/k. Once the $k+1^{st}$ participant is recruited, each member has a chance of 1/(k+1) of recruiting the $k+2^{nd}$ participant, etc. Thus, the number of people the k^{th} participant will recruit is expressible as the sum of independent binomial r.v.'s,

$$S_k = \sum_{i=k}^{N-1} X_i, \qquad (3.1)$$

where each

$$X_i = \begin{cases} 1, & \text{with probability } p_i = 1/i \\ 0, & \text{with probability } 1-1/i. \end{cases}$$

Thus, the expected number of people the k^{th} person will recruit equals

$$\sum_{i=k}^{N-1} 1/i \sim \ln[(N-\tfrac{1}{2}) / (k-\tfrac{1}{2})]. \qquad (3.2)$$

An immediate consequence of (3.2) is that once k is $\geq$ N/e, or about .37N, any future participant can expect to recruit no more than <u>one</u> person. Thus, only the 37% who join <u>first</u> can expect to recruit at least <u>one</u> new participant.

Another approach to demonstrating that a participant who joins the scheme after its initial phase has a <u>small</u> chance of recouping their investment is to calculate the probability that they will recruit the minimum number of people b = [c/d]+1, to achieve this. In our illustrative example, this value is 3. In order to compute $P(S_k \geq 3)$, statisticians use the Poisson approximation to the sum of binomials (3.1), as the p_i are small and decrease to zero. In the Appendix we describe a method of approximating S_k by Poisson r.v. P_k which is "stochastically larger" than S_k, and the probabilities presented in the table are derived from these results and are therefore upper bounds for the actual probabilities. (See table at the end of paper) The results in the table show that once a quota pyramid reaches <u>one-third</u> of its limit the probability a new member will regain his investment is less than 10%.

<u>The Expected Return to All Participants</u>

In the previous section we were concerned with the probability of each individual recruiting enough future members to regain the entrance fee. We now demonstrate that pyramid scheme investors are defrauded as a class.

The simplest proof of this is to notice that at any stage of the process (say K people are enrolled), the promoter (the first person) has received (K-1)c and has paid out (K-2)d. Hence, the promoter has a net profit of

$$c + (K-2)(c-d),$$

and the fraction of investment that has been returned to the participants is

$$\frac{K-2}{K-1}\,\frac{d}{c}\ .$$

Thus, the portion of all invested dollars returned to the participants is slightly less than d/c, the ratio of the fee earned for recruiting one new member to the initial investment. In the actual case used for illustration, this is only .36. Thus, as a class, participants will lose 64% of their investment.

Another interesting consequence of the

probability model is that on the average about half of the participants will recruit nobody and will lose their whole investment. This can be seen by noting that the probability that the k^{th} entrant will fail to recruit anyone is

$$P_k(0) = \prod_{i=k}^{N-1} (1-1/i) = (k-1)/(N-1).$$

Thus, the expected number of participants who are "shut out" is

$$\sum_2^N P_k(0) = \frac{1}{(N-1)} (1+\ldots+N-1) = \frac{N}{2},$$

i.e., half of the investors will lose everything they paid to join the system. Moreover, this remains true for any value of d (the amount paid for enrolling a new member). Thus, even if all the money paid in were returned to investors, half of them can expect to receive nothing.

One might question the relevance of the previous result in the context of a fraud case if a significant fraction of the participants were "big winners". When we replace the r.v. S_k denoting the number of people the k^{th} entrant recruits by its Poisson majorizer P_k, one can show (see Appendix) that as $N \to \infty$, the proportion of the participants who recruit exactly r people approaches $2^{-(r+1)}$ so that the fraction who recruit at least r is 2^{-r}. Thus only one-eighth of the participants can expect to recruit at least three members, and only one in 16 million can expect to recruit 24 or more people. Thus, our model agrees with the findings of Judge Naruk in the case described when he noted that no one had earned an amount of money near that claimed in the brochure.

In light of this and other facts, Judge Naruk permanently enjoined the defendants from selling or authorizing others to sell Goldren Book dealerships and from instituting any other multi-level merchandising plan in Connecticut without express court approval.

Appendix: A Probability Bound for the Sum of Poisson-Binomial Variates

Let X_i, $i=1,\ldots,n$, be independent binomial r.v.'s with $p_i=P(X_i=1)$ and let $S=\Sigma X_i$. When the probabilities p_i are "small", we desire a tight upper bound rather than an approximation to

$$P\{\sum_{i=1}^n X_i > a\}, \qquad (1)$$

where a is a specified integer usually greater than the expected value, Σp_i, of the r.v.'s.

In order to derive a bound for (1), we introduce Poisson r.v.'s Y_i, which are stochastically larger than the X_i's, i.e. we choose the parameter λ_i of Y_i to satisfy

$$P(Y_i=0) = P(X_i=0) = 1-p_i, \qquad (2)$$

i.e.,

$$e^{-\lambda_i} = 1-p_i \text{ or } \lambda_i = -\ln(1-p_i). \qquad (3)$$

In order to give X_i and Y_i a bona fide joint distribution, following Hodges and LeCam, we define

$$P(X_i=0,\ Y_i=0) = 1-p_i$$

and

$$P(X_i=1,\ Y_i=k) = \frac{e^{-\lambda_i}\lambda_i^k}{k!} \qquad (4)$$

where λ_i and p_i obey (3).

As $X_i \leq Y_i$ for each i, $\Sigma X_i \leq \Sigma Y_i$ and

$$P(\Sigma X_i \geq a) \leq P(\Sigma Y_i \geq a). \qquad (5)$$

As ΣY_i has a Poisson distribution, the probability on the right is readily computable once λ_i is expressed in terms of p_i. From the Taylor expansion,

$$-\ln(1-x) = \sum_1^\infty x^j/j,$$

it follows that

$$\sum_{j=1}^{k} \frac{x^j}{j} \leq -\ln(1-x) \leq \sum_{j=1}^{k-1} \frac{x^j}{j} + \frac{x^k}{k}\frac{1}{(1-x)} \qquad (6)$$

so each λ_i can be obtained to any desired accuracy. For practical purposes, the choice of k=3 usually suffices, so (6) becomes

$$p_i + \frac{p_i^2}{2} + \frac{p_i^3}{3} \leq \lambda_i \leq p_i + \frac{p_i^2}{2} + \frac{p_i^3}{3}\frac{1}{(1-p_i)}.$$

When the $\{p_i\}$ decrease, the difference between the bounds on the parameter

$$\sum_j^N \lambda_i$$

of the Poisson r.v. majorizing S is

$$[(1-p_j)^{-1}-1][\sum_{i=j}^N p_i^3]/3.$$

Before applying the above method to our special case we present the analog of Hodges and LeCam's results for the difference between $P(S \geq a)$ and our approximation P_λ. Specifically, we have

Lemma: For any constant a,

$$P(P_\lambda \geq a) - P(S \geq a) \leq \Sigma p_i^2 \qquad (7)$$

Proof: For each i,

$$P(Y_i > X_i) = P(Y_i \neq X_i) = \sum_{k=2}^{\infty} e^{-\lambda_i} \frac{\lambda_i^k}{k!} =$$

$$1 - e^{-\lambda_i} - \lambda_i e^{-\lambda_i} = p_i + (1-p_i)\ln(1-p_i).$$

As $e^{-x} \geq 1-x$, $-x \geq \ln(1-x)$, so

$$p_i + (1-p_i)\ln(1-p_i) \leq p_i - (1-p_i)p_i = p_i^2.$$

By Boole's inequality, $P(P_\lambda > S) \leq \Sigma p_i^2$.

Application to the Pyramid Scheme

In our example, $p_i = 1/i$, and we desire to approximate

$$S_k = \sum_{i=k}^{N=1} X_i$$

by a Poisson variable.

In our case we can obtain an explicit expression for γ_k rather than using a Taylor series development as

$$\lambda_i = -\ln(1-1/i) = \ln(i/i-1).$$

Thus,

$$\gamma_k = \sum_{i=k}^{N-1} \lambda_i = \sum_{k}^{N-1} [\ln i - \ln(i-1)] = \ln((N-1)/(k-1))$$

so that

$$P(S_k \geq r) \leq \sum_{i=r}^{\infty} e^{-\gamma_k} \frac{\gamma_k^i}{i!} = 1 - \sum_{i=1}^{r-1} \frac{(k-1)}{(N-1)} \frac{\gamma_k^i}{i!} . \qquad (8)$$

Since S_k is the sum of non-identically distributed binomial r.v.'s, a compact formula for its exact distribution is not available and a computer is needed. In Table A.1, we compare the exact value of $P(S_k \geq 2)$ and $P(S_k \geq 3)$ to the bounds we obtained from formula (8). Clearly the bounds are quite close.

As our r.v.'s P_k approximating S_k are so close we can derive an accurate approximation to the expected fraction of participants who will recruit at least r people. Formally, we have

Theorem: Let $X_2, X_3, \ldots, X_N$, be a sequence of Poisson r.v.'s with parameters

$$\gamma_k = [\ln(\frac{N-1}{k-1})].$$

Then

$$\frac{1}{(N-1)} \sum_{k=2}^{N-1} P(X_k = r) \to 2^{-(r+1)}, \quad r=0,1,2,\ldots \qquad (9)$$

as $N \to \infty$.

Proof: As $P(X_k=r) = \gamma_k^r e^{-\gamma_k}/r!$, the left side of (9) is

$$\frac{1}{r!} (N-1)^{-1} \sum_{2}^{N-1} \frac{(k-1)}{(N-1)} [\ln(\frac{N-1}{k-1})]^r . \qquad (10)$$

Letting $v = (k-1)/(N-1)$, (10) is a Riemann approximation to

$$(r!)^{-1} \int_0^1 v[\ln(\frac{1}{v})]^r dv =$$

$$(r!)^{-1} \int_0^\infty z^r e^{-2z} dz = 2^{-(r+1)}.$$

Hence, for large N, the expected fraction of all participants who recruit at least r people is $1/2^r$ for r=0,1,2,...

In order to see how fast the limit is approached we computed the exact values of (9) when N=270 and 1000 for r=1,2, and 3. The resulting values which have limits 1/4, 1/8, 1/16, were .24991, .12475 and .06202 (N=270) and .24999, .12497, .06244 (N=1000)

References

[1] Anderson, T.W., and Samuels, S.M. (1965) Some Inequalities Among Binomial and Poisson Probabilities. Fifth Berkeley Symposium. Vol. I. pp. 1-12.

[2] Feore, J. (1974) Pyramid Scheme: Dare to be Regulated. Georgetown Law Journal. pp. 1257-1293.

[3] Ger-ro-Mar. Inc. vs. F.T.C. (1975) 518 F 2d 33.

[4] Hodges, J.L., Jr., and LeCam, L. (1960) The Poisson Approximation to the Poisson Binomial Distribution. Ann. Math. Statistics. Vol. 31, No. 3 pp. 737-740.

[5] Hoeffding, W. (1956) On the Distribution of the Number of Successes in Independent Trials. Ann. Math Statistics. Vol. 36. pp. 713-721.

[6] Naruk, H. (1975) Memorandum of Decision: State of Connecticut vs. Bull Investment Group. 32 Conn, Sup. 279.

This research was supported in part, by National Science Foundation Grant MPS73-08444.

Addenda: The Case in Which the Court Rejected the "Geometric Progression" Argument

In order to motivate the development of our model, this addenda will describe the decision in the Ger-Ro-Mar vs. F.T.C. case [3]. The company manufactured and sold brassieres, girdles, swimwear and lingerie under the label Symbra'Ette. Its sales grew from $37,000 in 1965 to $2 million in 1969 but fell to $1.2 million in 1972. The company sold its products through a sales force which required distributors to buy an inventory of products in order to participate. The sales organization was a multi-level one in which supervisors earned a percentage of the sales of those below them. The entry level (Key Distributor) required a purchase of $300 retail value of merchandise (i.e. an initial investment of $215, the wholesale value). A Key Distributor sold the products door to door and could also engage in unlimited recruiting and become a Senior Key Distributor when the retail value of the merchandise he and his recruits reached $1000 in any month. The profit for Senior Keys consisted of an increased profit margin on their own retail sales and a percentage profit on the purchases made by his recruits and various related commissions. Similarly, Senior Keys could rise to higher levels when they and their recruits achieved the requisite retail value of products purchased from the company.

To induce individuals to participate in the program the promotional brochure illustrated how, both by building a large personal group of sales people via recruitment and by selling at retail, a person could earn large sums of money, e.g. $56,400 per year as a District Manager.

Before quoting the 6th Circuit's decision, it should be noted although there is a pyramiding aspect in this program, the situation differs from the Golden Book case because

1) the initial investment was relatively small ($215) and

2) since there was a product available to sell one did not have to recruit others to rise in the "system" in order to earn money.

We now quote from the decision:

"The sole evidence to support the Commission's holding that the plan is inherently unfair and deceptive is a mathematical formula, which shows that if each participant in the plan recruited only five new recruits each month and each of those in turn recruited five additional recruits in the following month, and this process were allowed to continue, at the end of only 12 months the number of participants would exceed 244 million including presumably the entire staff of the FTC. The Commission concludes that this, in effect, is the impossible dream and that the siren song of Symbra'Ette must be stilled. We find no flaw in the mathematics or the extrapolation and agree that the prospect of a quarter of a billion brassiere and girdle hawkers is not only impossible but frightening to contemplate, particularly since it is in excess of the present population of the Nation, only about half of whom hopefully are prospective lingerie consumers. However, we live in a real world and not fantasyland.

As indicated by the record, the fact is that Symbra'Ette, which commenced business in 1963, did not reach its peak in distributorships until 1972 when it had attracted some 3,635 distributors. The record does not indicate the geographical distribution of these vendors, and we have no study or analysis in the record which would realistically establish that some recruiting saturation exists which would make the entry of additonal distributors and the recruitment of others potentially impossible in any practical sense. While the Commission need not establish actual deception by the testimony of disappointed entrepreneurs, it has failed entirely to establish a potential threat. Not all Americans aspire to the calling in issue and not all who are attracted will continue indefinitely."

Apparently the F.T.C. relied solely on the "geometric progression" argument rather than obtaining data to estimate the crucial statistical quantities such as the average number of new recruits each participant achieves or the proportion of sales persons who recruit no one else. Without such supporting evidence, it is difficult to convince a Court that it is mathematically impossible for a business to survive when it has existed for a number of years. Although the author hasn't seen the company's books, he feels that they would show that a substantial portion of the merchandise sold was to new recruits rather than to the public at retail.

Table 1 The Expected Number of People Each Participant will Recruit and Upper Bounds for the Probability of Recruiting at Least 2 or 3 New Members (N = 270)

Position of Entry	Expected No. of Recruits	Probability of Recruiting at Least r New Members r=2	r=3
k= 5	4.208	.9226	.7909
10	3.398	.8529	.6598
20	2.6500	.7422	.4941
30	2.227	.6521	.3846
40	1.931	.5750	.3047
50	1.703	.5077	.2435
60	1.517	.4479	.1955
75	1.291	.3699	.1407
90	1.106	.3032	.1008
100	1.1000	.2641	.0802
120	.8160	.1968	.0497
135	.697	.1547	.0338
150	.591	.1189	.0222
180	.407	.0635	.0083
210	.2524	.0270	.0022
240	.1182	.0065	.0003

Table A.1 Comparison of Our Bounds to the Exact Values (N=270)

Index (k)	Exact $P(S_k \geq 2)$	UB for $P(S_k \geq 2)$	Exact $P(S_k \geq 3)$	UB for $P(S_k \geq 3)$
5	.92064	.92255	.78289	.79087
10	.85104	.85287	.65358	.65978
25	.69346	.69517	.43053	.43464
40	.57344	.57504	.30162	.30470
50	.50613	.50765	.24095	.24354
75	.36851	.36986	.13900	.14074
90	.30195	.30319	.09943	.10081
135	.15378	.15472	.03311	.03376
180	.06290	.06353	.00807	.00833
210	.02655	.02696	.00212	.00222
240	.00626	.00646	.00023	.00025

Remark: The ordinary Poisson approximation also yields precise estimates of the exact probabilities. These estimates are slightly low for the early values of k and become slightly high for larger values of k as would be expected from the literature [1] on the Poisson approximation to the sum of i.i.d. binomial r.v.'s.

DISCUSSION

William B. Fairley, Commonwealth of Massachusetts

The paper by Haworth and Haworth puts a useful emphasis on the use of implicit models by the parties or by the judge in legal cases. For example, in the jury discrimination cases an implicit model of random selection from a pool of individuals qualified for jury duty is assumed. In wage discrimination cases an implicit model of wage determination assumes that wages are paid on the basis of factors that are distributed alike in all groups. One of the contributions of statistical analysis is to lay bare implicit models and deduce some of the implications that flow from models. In the jury cases a random model of fair selection of jurors may be the appropriate legal standard and implications drawn from it by mathematical and statistical reasoning may assist the judge in making a correct determination under the given standard (See Finkelstein, "The Application of Statistical Decision Theory to the Jury Discrimination Cases", 80 Harvard Law Review 338 (1966).)

An implicit model usually underlies a quotation of the probability of some event. In one of the jury discrimination cases referred to in the paper the percentage of Blacks on juries had not exceeded 15 percent for 15 years though they comprised 26 percent of the population, and the probability of this event was quoted as 4.63×10^{-21}. This small probability was computed under a model of independent random selections for the jurors in a series of juries in 15 years. The value of the probability is of course very much dependent on the model assumed, and this point is often lost when the value is produced as the probability without reference to the model. As a footnote to this example I would add a reference to the discussion of very small probabilities, such as the one quoted, by Mosteller and Wallace in Inference and Disputed Authorship: The Federalist (Addison-Wesley, 1964). They discuss the inherent incredibility of a very small probability of an event whenever there are "outrageous events" that might produce the given event which themselves have probabilities larger than that of the small probability quoted. The result is a *reductio ad absurdum*. The small probability cannot be that small.

Brosi and Brounstein present part of a larger and continuing research effort. Their work touches on a number of interesting questions for people interested in how statistics can be used to aid decision making in the law. Are there statistically derived functions available about a defendant that can be used to improve decision making about bail and other matters? If so, where and how is it appropriate for the judge to enter his subjective opinions? Work in the psychological literature on these questions was stimulated some time ago by Paul Meehl in Clinical versus Statistical Predictions (University of Minnesota Press, 1954). See more recent discussions, for example, Robyn Dawes, "A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making", American Psychologist, February 1971. Current literatures on predicting parole success and on predictive efforts in preventive detention and in other legal contexts are also relevant.

Could the research discussed in the Brosi and Brounstein paper relate to legal and policy issues surrounding preventive detention and bail? What might the research have to say about the difficulties of prediction that judges face? What are their likely Type I and Type II error rates? Will anything be said about differences between judges?

The use of some of the information on defendants that is discussed in the paper is circumscribed by law or custom. For example, prior criminal record and the sex or race of the defendant might or might not be permitted to enter as a "variable" in some of the uses of the research that are proposed. How are these issues to be handled?

The Gastwirth paper raises the following issue. Legal standards, though often stated informally in terms of chances, are rarely defined with reference to precisely stated probabilistic criteria. The paper shows in several precise ways how poor a lottery it was that Glen Turner offered to prospective dealers. How were standards of fraud (and other appropriate standards if any)applied in this case?

The probabilities computed in the paper rest on the assumptions of the model, in particular that the probability that any specific one of k current members of the scheme recruits the next member is 1/k. Suppose this assumption were relaxed, for example, by recognizing differences in ability among members. What would be the effect on the probabilities, which would now differ by member, and how would this affect the finding of fraud?

The use of an upper bound for the Posson approximation to the sum of binomials was a clever way to avoid legal challenge. We may still ask whether a suitably stated approximation not involving an upper bound would have served.

SESSION ON GRAPHICAL METHODS FOR PRESENTING STATISTICAL DATA

R. Bachi, Hebrew University, Jerusalem, Israel

Statistical graphics is passing today through a phase of its development which is at the same time challenging and critical.

Automation has opened possibilities which a few years ago would have appeared as Utopian. For instance:

a) Graphs can today be mass produced at a tremendous speed and low cost;

b) Automated graphs can attain levels of accuracy and attractiveness even exceeding those of hand made graphs;

c) Both effectiveness and beauty of the automated graph can be enhanced with colors;

d) In each stage of graph production, quick and efficient experimentation of alternative solutions, and substitution and alteration of the various elements of the graph become possible.

e) These facilities, together with animation, give to the graph powerful potentialities as scientific research tools and as suggestive means of communication.

In the second paper of this session, Mr. Barabba will introduce us to the wonderland of graphical automatic devices by showing the amazing example of the graphical activities which have been developed in the course of a few years at the Bureau of the Census.

The technical revolution in the preparation of graphs has taken place in a period in which demand for graphs is increasing at a very considerable pace. Both the papers by Dr. Schmid and Mr. Barabba illustrate how the modern explosion in data production entails an increased need of graphs for spreading part of this information among the general public, and for providing a tool of quick communication of essential knowledge to decision makers, both in politics and business. It is likely, too, that the importance of the graphs, as instruments of scientific and practical research, may grow in the future along with increasing awareness of how the graphs can help us to obtain an overview of the data, to discover regularities and irregularities, to suggest new scientific hypotheses and to discard previous ones in the light of empirical evidence. A further task which may become very important is that of translating the enormous masses of data produced by statistical offices into a more understandable form than the tabular one and to help in deciding what sets of data are worthy of publication.

The coincidental increase in technical possibilities and in demand for graphs is probably not fortuitous and is certainly welcome. However, it is both paradoxical and sad that those developments have not been accompanied by a corresponding development of graphical statistical methodology and of its theoretical background.

As will be indicated in the paper by Dr. Schmid, the situation of these fields has remained stagnant until very recently. Graphical methodology has largely disappeared from most Universities' curriculum as a subject worthy of extensive and deep didactic and research activity. Findings of psychological research of the type which will be illustrated by Dr. Wainer have not been properly channeled to the knowledge of professional statisticians and even less to that of graph makers. The same applies also to findings of information sciences.

The retreat of statisticians from the field, well illustrated by Dr. Schmid, has left a vacuum which is being filled more and more by technicians, whose first interest is rationality of production and not fidelity of the graph to the data represented. Therefore, a sizeable part of the graphs produced today are of poor scientific quality, may unintentionally deceive the public and the decision makers for whom they are intended and, in the long run, they may bring discredit to statistical graphics.

Some awareness of those dangers seem to have emerged in the past few years. It may perhaps not be accidental that in the last few years some professional statisticians have taken an interest in the field, graphical methods have been debated at national and international statistical conferences, some professional groups of users of graphs have been formed and the Bureau of the Census has taken a lead in this field. Also our meeting of today is, to some extent, a proof of this awareness.

The program of our meeting, as prepared by Mr. Barabba, is a well balanced one as it gives to us many facets of the problems ahead of us.

The paper by Mr. Barabba will indicate what technology has done and still can do to improve the field, and how a big data producer like the Bureau of the Census can shift more and more toward graphs as means of communicating its findings.

The paper by Dr. Wainer will suggest avenues in which psychologists can help the statisticians in the preparation of graphs apt to give the information they are purported to give.

The Poster Session, immediately following this meeting, will show examples of developments of graphical methods which can be achieved (i) by introducing new graphical symbols or tools; (ii) by clarifying some basic concepts in certain fields of graphical presentation; and (iii) by eliminating limitations and shortcomings of classical methods.

The paper by Dr. Schmid will introduce us to the important problem of role of standards in graphical presentation.

I take the liberty to suggest that in the discussion of this valuable paper, attention be paid also to the hints contained in it on practical ways which can be found to ensure some follow-up to the meeting of today. It seems highly desirable, that some channels may be established at national and international levels, to render more efficient the work needed to reassess the entire graphical field. As a final aim it seems desirable to enlarge, improve and revise the graphical methodology, to formulate better standards, to enlarge the attention given to graphics in university curricula, to train specialists and to give a better education of statistical requirements to graph makers.

THE ROLE OF STANDARDS IN GRAPHIC PRESENTATION

Calvin F. Schmid, University of Washington*

Standards are a universal ingredient of every type of orderly human relationship whether it be in government, business, industry, religion, science or in any other area or activity. Since "standard" and "standardization," two widely misunderstood and misused concepts, will be cited repeatedly in this discussion, it is essential at the very beginning to define them with some degree of clarity and specificity. A "standard" is a criterion, unit of reference, model or process approved or accepted as correct by common consent, established custom, or recognized authority. Frequently, standards evolve without conscious direction through such processes as common practice, imitation, and precedent. Also, standards may be formalized and systematized through consensus by special committees or groups created for such purpose. Standards exist in different forms such as (1) a document or systematized formulation containing a set of conditions to be fulfilled in accordance with specified rules and directions; (2) a fundamental unit or physical constant (examples: ampere, absolute zero); (3) an object for physical comparison (examples: meter, liter). "Standardization" is the process of formulating and applying rules for an orderly approach to a specific activity. Standardization is not a series of mandatory edicts; it is not a strait jacket of conformity; nor is it an exposition of dull, drab rules. Standardization means consensus and cooperation for the purpose of attaining optimal economy and efficiency. It is a form of conscious planning based on the consolidated results of science, techniques and experience. Some particular applications include: (1) units of measurement; (2) terminology and symbolic representation; and (3) rules and instructions pertaining to products and processes.[1]

With respect to the origin of standards, a general distinction can be made between those that are based on habit, custom or tradition which can be designated "natural standards," and those that are the result of conscious planning which can be designated "organized standards."[2] Certainly in graphic presentation as in other fields both "natural" and "organized" standards will be found.

For those of you who still retain at least a distant memory of introductory sociology will recognize that "natural standards" possess a meaningful similarity if not identity to such concepts as folkways, customs, mores, norms, and other elements of normative systems in human society. In fact, they may be referred to properly as standards of behavior. Characteristically, these elements and patterns develop spontaneously and unconsciously and serve as standards and guides to human conduct. As a society grows and becomes more complex the "natural" patterns and standards based on tradition and experience evolve into formal prescriptions and laws. This change exemplifies the transformation of "natural" standards into "organized" standards.

The following example will provide a historical glance of this process, as illustrated by the transformation of standards of linear measurement from crude and informal beginnings to more objective and precise criteria. Many centuries ago

> . . . it was sufficient that various parts of the human body serve as measuring units since they were handy and required no unusual skill to use. For instance, one of the earliest standards of measurement was the cubit, which was the length of the forearm from the point of the elbow to the tip of the middle finger. Later the inch was the width of a man's thumb; the foot was the length of the reigning king's foot; and the yard was the distance from the thumb to the tip of the nose. During one period the standard for the inch even became the length of three pieces of barleycorn from the "middle of the ear."
>
> In time, with the increase in commerce and communication, it became obvious that units of measurement could not be based on variables such as thumbs, elbows, noses and corn.
>
> . . . The French revolution not only brought drastic social and political innovations, but also gave birth to the metric system. This introduces a comprehensive decimal system having as a basis the meter, which was taken as the one ten-millionth part of a meridional quadrant of the earth.[3]

It is significant to observe that it is not uncommon for widespread resistance to develop against the adoption of new and demonstrably superior standards. As you know, many years passed before the metric system was adopted as the obligatory system in France and other countries. As far as the United States is concerned a Congressional act was passed in 1866 making it "lawful throughout the United States of America" to employ the system and defining meter in terms of inches. In recent years much is heard about the adoption of the metric system but it will be several decades before any substantial transition to the metric system is achieved. Such factors as cost, confusion and general cultural inertia preclude any rapid changeover of this kind.

Industrial Standards and Standardization and Their Influence on Standards of Graphic Presentation

A preliminary discussion of standards and standardization would be seriously deficient if at least brief reference were not made of the impressive role of industrial standardization, both nationally and internationally. Because of its pervasive influence on every facet of our economy its true significance and impact on modern technological civilization is not fully grasped. However, it can be said that "The partnership between science and standards holds the secret to the extraordinary dynamism and productivity of modern industrial technology."[4] "Without standards, our present-day economy would be a shambles --in fact, it might never have come into being."[5]

For several decades thousands of private organizations and individuals along with numerous governmental agencies have been actively involved in the industrial standardization movement. Many professional organizations and agencies such as the American National Standards Institute, the American Society of Mechanical Engineers and the United States Bureau of Standards have played key roles.

The industrial standardization movement by providing a pervasive stimulus exerted an influence in the origin and development of standards of graphic presentation. Willard C. Brinton, a professional engineer, through the American Society of Mechanical Engineers was largely responsible for the original Joint Committee on Standards of Graphic Presentation. Subsequent committees that revised and expanded the original standards were sponsored by the American Society of Mechanical Engineers functioning under procedures and requirements of the American National Standards Institute.[6]

Brief Historical Background of Standards of Graphic Presentation

In an effort to acquire a better understanding of the significance and purpose of standards in graphic presentation, an historical perspective will be found particularly helpful. Accordingly, consideration will be given to a brief historical account of graphic presentation, highlighted by certain facts pertaining to standards and standardization.

The origin of statistical charting techniques, as we think of them today, dates back to 1786--less than 200 years--when William Playfair published his famous work entitled The Commerical and Political Atlas. Two subsequent editions of this book were published in 1787 and 1801. It must be recognized, of course, that in studying the history of graphic techniques many basic developments such as the principle of coordinates and the invention of analytic geometry, antedate the work of Playfair. Nevertheless, William Playfair can properly be considered the "father" of graphic presentation.

Playfair, in refering to his new system as "lineal arithmetic" explains that

> The advantage proposed by this method, is not that of giving a more accurate statement than by figures, but it is to give a more single and permanent idea of the gradual progress and comparative amounts, at different periods, by presenting to the eye a figure, the proportions of which correspond with the amount of the sums intended to be expressed.[7]

Furthermore, he states

> That I have succeeded in proposing and putting in practice a new and useful mode of stating accounts, has been so generally acknowledged, . . . as much information may be obtained in five minutes as would require whole days to imprint on the memory, in a lasting manner, by a table of figures.[8]

It must not be overlooked that when Playfair published his original contributions, including the line graph, circle graph, bar graph and pie diagram the word "statistics" had not yet appeared in the English language, few collections of reliable quantitative data were available, and the development of statistical method was still far in the future.[9]

Standards and standardization are nothing new in graphic presentation. When Funkhouser wrote his well-known history of graphic presentation in 1937, he stated that "the problems met in trying to classify and standardize graphic forms have been wrestled with for almost a hundred years."[10] During the early history of graphic presentation, the significance of these problems as well as the manner in which they were considered are reflected in the proceedings of the nine International Statistical Congresses that were held in Europe from 1853 to 1876. Subsequently, the International Statistical Institute which was organized in 1885 gave serious consideration to standards of graphic presentation. Graphic techniques were discussed at considerable length at the International Statistical Congress held in Vienna in 1857. For many years, concomitant with the rapid growth of graphic presentation, many statisticians believed that an effort should be made to regulate graphic procedures and to provide rules for the purpose of achieving uniformity and comparability. In the Hague Congress of 1869 a resolution was passed recommending that the "organizing commission of the next Congress prepare a memoir on the different graphic methods employed in statistics and on the proper means of rendering the graphic tables uniform and comparable."[11] Accordingly, in the ensuing Congress held at St. Petersburg in 1872, this issue was faced head on. After extended discussions and debates before the general assembly it was concluded that "As for uniformity of diagrams, properly called, the Congress declares that the time has not yet come to prepare uniform rules."

> The issue on uniformity of graphic procedures was not drawn as sharply again. Most statisticians came to realize that, although some standard practice in the drawing of diagrams was desirable, the type of regulation urged at the St. Petersburg Congress was both foolish and impractical. For the next forty years the matter was debated informally at statistical gatherings and by various writers but nothing constructive was accomplished.[12]

Following the last statistical Congress in 1876 in Budapest, and an ineffectual attempt to hold an assembly in Rome in 1880, the International Statistical Institute was organized at the jubilee of the London Statistical Society in 1885. At the sessions of the International Statistical Institute in 1901, 1908, and 1913 serious attempts were made to develop rules and standards for graphic procedures, but they all met with failure.

American Joint Committee on Standards for Graphic Presentation

Meantime in 1914 in the United States, largely through the efforts of Willard C. Brinton, the American Society of Mechanical Engineers extended invitations to a number of interested American scientific societies to participate in a joint committee for the purpose of developing standards of graphic representation. Seventeen associations and agencies cooperated in the formation of the committee. The initial meeting of the committee was held in December 1914. The first report, described as preliminary, was published in 1915.[13] The report was relatively brief, consisting of 17 simply stated basic rules, each illustrated with from one to three diagrams. Fourteen of the rules including the accompanying diagrams were devoted exclusively to the portrayal of time series in the form of arithmetic line charts. Of the three remaining rules, one emphasized the preference of linear magnitudes over areas or volumes; one represented a simple procedure pertaining to semi-logarithmetic charts; and one, the desirability of emphasizing the 100 percent or other base line in the delineation of an arithmetic grid.

Since the publication in 1915 of the report by the original Joint Committee, other committees prepared greatly expanded reports on standards of graphic presentation in 1936, 1938, and 1960.

Present-Day American National Standards Committee on Preferred Practice for the Preparation of Graphs, Charts and Other Technical Illustrations

At the present time, there is a permanent committee on standards of graphic presentation, officially known as Y15 American National Standards Committee on Preferred Practice for the Preparation of Graphs, Charts, and Other Technical Illustrations. It was organized in 1926 and re-organized in 1949. There are many American National Standards Committees, mostly in business and industry, that have been organized and are functioning under the auspices and in accordance with certain rules and specifications of the American Standards Institute. The sponsor for Y15 American National Standards Committee on Preferred Practice for the Preparation of Graphs, Charts and Other Technical Illustrations is the American Society of Mechanical Engineers.

Because of limitation of space, it will not be possible to present a detailed explanation of the methods and sanctions specified by the American Standards Institute in establishing an American National Standards Committee. However, an attempt will be made to provide a meaningful sense of some of the more basic requirements and procedures especially as they apply to the American National Standards Committee on Preferred Practice for the Preparation of Graphs, Charts, and Other Technical Illustrations.[14]

1) The American National Standards Institute shall consider any written request to establish an American National Standards Committee.

2) Such requests shall include the (a) proposed scope of the committee, (b) a history of standardization work in this field and (c) a list of organizations having a substantial concern with, and competence in, the proposed scope.

The official scope of ANSC on Preferred Practice for the Preparation of Graphs . . . is as follows: "The recommendation of preferred practices for the design and preparation of graphs, charts, and other technical illustrations, including consideration of special requirements for publication or projection."

3) Every ANSC is required to have a secretariat (sponsoring organization) that is charged with certain specified functions and responsibilities, such as: (a) carrying out the Institute's procedures for the ANSC; (b) determine representatives on the ANSC; (c) propose programs of work, together with proposed completion dates; give direction and guidance to the ANSC; and (d) carry out administrative work, including secretarial service.

The secretariat for the ANSC on Preferred Practice for the Preparation of Graphs . . . is The American Society of Mechanical Engineers.

4) Membership on an ANSC is of three types: (a) representatives of organizations "willing to participate and having substantial concern and competence in the scope of the Committee"; (b) "individuals possessing expert knowledge in the field of the Committee's work"; and (c) under certain conditions "companies having substantial concern and competence in standards within the Committee's scope."

The present ANSC on Preferred Practice for the Preparation of Graphs . . . is composed of 16 members: 12 representing professional and trade associations; 3 individual members; and 1 industrial member ("telephone group"). In addition there are 2 alternate members. Significantly, and strangely, the American Statistical Association is not represented on this Committee. Moreover, according to the 1970 directory, not a single one of the 16 members and 2 alternate members of the Committee are members of the American Statistical Association. In contrast, among the 17 members of the original Joint Committee in 1915, the American Statistical Association was represented by Leonard P. Ayres who was elected secretary of the Committee. The chairman of the Committee, Willard C. Brinton, was a former vice president of the American Statistical Association. In addition there were other well known statisticians such as Robert E. Chaddock, Edward L. Thorndike, and Joseph A. Hill who served on the Committee. In the 1930's, when the 1936 issue of Code of Preferred Practice for Graphic Presentation--Time-Series Charts and the extensive 1938 revision entitled Time-Series Charts--A Manual of Design and Construction were published, the American Statistical Association was represented by Karl G. Karsten. In the 1960 revision, the professional society affiliation of the members of the subcommittee responsible for this report is not indicated. However, three of the nine members are members of the American Statistical Association.

Original American Standards as Well as Subsequent Revisions Devoted Exclusively to Time-Series Charts

The standards of graphic presentation published originally in 1915 along with the extensive revisions and additions promulgated in 1936, 1938 and 1960 are excellent examples of rationally organized and formalized standards. Without seeming repetitious, it should be emphasized that for the most part these standards existed long before 1915. They evolved over the years through practice, experience, imitation and precedent. Those responsible for the published formalized standards were basically codifiers and organizers who selected and refined certain rules, procedures and practices through discussion, evaluation and consensus. A careful examination of the published standards of graphic presentation, valuable as they have been, are limited in application since they are concerned exclusively with time-series charts.

For example, in the 1938 edition of Time-Series Charts: A Manual of Design and Construction over 50 pages are devoted to arithmetic line charts, 3 pages to time-series column charts and 2 pages each to surface charts and semi-logarithmetic charts.

Apparently, the rationale for selecting time-series charts as a basis for formulating standards was their extensive use and widespread familiarity. According to the 1938 edition of the Manual:

> Probably three quarters of all charts prepared employ time as one of the variables. Of the various types of time-series charts, the socalled "line chart" is most frequently used, and therefore is given the most space and is discussed in the most detail.[15]

The committee that prepared the 1960 revision continued the emphasis on time-series charts.

> The objective of the original committee who prepared the earlier version of this manual in 1938 was to bring together the principles and procedures found successful in constructing time-series charts. The objective of the present committee has been to review these principles and to revise the procedures to agree with current practices.
>
> In the years since the original manual was prepared, many of the practices used in the preparation of time-series charts have changed.[16]

In the 1960 revision 62 pages are devoted to time-series arithmetic line charts, 5 pages to time-series column charts, 3 pages to time-series surface charts and 6 pages to time-series semi-logarithmetic or ratio charts.[17]

It would be difficult to fault the committees (1914, 1936, 1938 and 1960) for the selection of time-series charts as the graphic form as a basis in the formulation of standards of graphic presentation. They possess a long tradition, are familiar to most people and are extensively used. However, during the past few decades other graphic forms have assumed increasing importance. In order to determine with a reasonable degree of reliability, the frequency of use of various types of charts, a survey far beyond the scope of the present paper would be required. The extent to which the various graphic forms are used is based on such general and specific factors as the following: field of study, characteristics of the data, cost, audience to whom the study is addressed, objectives of the study, and knowledge and expertise of the author.

For the purpose of deriving clues to the comparative extent to which certain types of graphic forms are utilized at the present time, let us consider a few recent studies. All of the studies are in the social sciences. The first is Social Indicators 1973 with more than 165 charts.[18] This monograph has received considerable attention from various groups as well as from individual scholars. Although, according to the introduction of this report "the indicators presented . . . are primarily time series," only 84, or 50.9 percent of the total of 165 charts can be classified as time-series charts. Specifically, 76 are arithmetic line charts, 6 are semi-logarithmic and 2 are surface or stratum charts. The remaining 81, or 49.1 percent consist of 73 bar and column charts, 5 histograms, 2 maps and one age-and-sex pyramid.[19] Three of our own recent studies, one devoted to a statistical and ecological study of Crime and two to demographic and ecological studies of Nonwhite Races and The Growth of Towns and Cities include a total of 271 charts.[20] Of the total of 271 charts only 7, or 2.6 per cent are arithmetic line charts. There are 11 other arithmetic time-series charts--10 surface or stratum charts and one column chart. In addition, there are 47 semi-logarithmic charts devoted to time series. The 18 arithmetic and 47 semi-logarithmic time-series charts comprise 24.0 percent of the total in the three monographs. The remaining 206 charts include cross-hatched maps, different kinds of spot maps, maps with 2- and 3-dimensional symbols, maps in perspective and in oblique projection, frequency polygons, bar graphs, correlations matrices, age-and-sex pyramids, organizational and flow charts and profile charts.

Non-Time-Series Charts and Standards of Graphic Presentation

Regardless of the precise percentage, there is no doubt that at the present time graphic forms other than those devoted to time-series comprise a very large proportion of charts in the armamentarium of the graphic specialist. Logically, this raises two significant questions with respect to standards of graphic presentation: First, what standards if any, exist for graphic forms other than time-series charts? and second, has the time arrived when a concerted effort should be made to formulate standards for at least some of the non-time-series charts?[21]

Apropos to the first question, standards for all the manifold graphic forms not classified as time-series charts do exist although they have not been explicitly organized and sanctioned through collective action by a special committee or organization. These standards are very real and meaningful and are an integral part of the discipline; they give direction to basic criteria, practices and techniques. They are commonly embodied in text books and manuals on graphic presentation. The value and utility of the standards thus presented depend upon the fidelity with which the standards conform to the best state of knowledge relating to the theory and practice of graphic presentation.

Because of the complex implications of the second question, "has the time arrived when formalized standards for certain non-time series charts to be promulgated?" a more than simple categorical yes-or-no answer is required. I believe that an appropriate and realistic answer can be achieved only after careful study by a committee of experienced and knowledgeable specialists from a number of relevant disciplines.

When graphic presentation is properly thought of as a graphic language, a form of visual communication, it can be readily seen how standards in graphic presentation are analogous to rules of grammar in the spoken and written language. As standards become more explicit and formalized through rational evaluation and consensus, graphic presentation can rid itself more easily of provincialisms, uncertainties, eccentricities and inconsistencies.

Standards should never be treated as ultimates. Sound standards of graphic presentation embody the best current usage, and are based on "general agreements" rather than on "scientific test." In the future perhaps, certain aspects of "general agreement" can be substantiated by "scientific test." Standards define knowledge at a point in time, usually by stating what is "best" when judged by some set of criteria. When knowledge increases or criteria change, standards must and do change. As experience in the field of graphic presentation broadens and deepens, and as new problems occur, changing practices are inevitable. New standards are created, and other standards become outmoded.[22]

Charts in General Publications as De Facto Standards of Graphic Presentation

Published charts, whether "good" or "bad" may have an impact not unlike existing standards since they are sometimes unconsciously imitated or used as models by those designing charts. Also, innovations and precedents may be established by this process. The influence of a publication may be particularly significant if it is prestigious and widely read. For example, during the more than 50 years since its publication, the 154-page volume by Leonard Ayres which was devoted to certain military aspects of World War I has been cited a number of times for the exemplary quality of its charts.[23] For example, one writer states that "It is probably one of the best graphic works done in this country up to that time."[24] Another writer indicates that this volume "contains some of the best graphic work done in the United States."[25]

On the other hand, a volume containing a large number of charts of poor quality may have an opposite influence. It is a disservice to the reader and to the discipline to publish poorly designed and executed charts. The graphic material, comprising 165 charts in a widely recognized volume--Social Indicators, 1973--which was previously cited in this paper is a case in point. For example, all of the 76 arithmetic line charts are aberrant and idiosyncratic in design and most of the remaining 89 charts are mediocre or actually violate accepted standards of graphic presentation.

In this connection it is significant to note that the Social Science Research Council sponsored a review symposium of this volume in which 37 statisticians and social scientists participated. A monograph comprising the basic proceedings of this symposium was published in 1974.[26]

In spite of the fact that the volume under review is referred to in the introduction of the Symposium as a "chartbook," and that well over half of the space in the volume under review is devoted to charts, there is virtually nothing in the Symposium on an evaluation of the graphic material. The symposium does include a cursory discussion of four charts in the context of specific statistical problems along with the puzzling statement that "the graphics are among the best we have seen in such a report, not only because of the helpful use of color, but also because the authors have generally observed relatively high standards of presentation."[27]

Role and Importance of Internal Standards

Although an extensive body of general standards, national or even international in scope, may exist for a particular discipline or other established area, it is a common expendient for constituent organizations or other entities to modify or supplement existing standards for the purpose of fulfilling their own special needs. In fact, in industry and business there are very few large organizations that have not established their own files of internal or in-house standards. Customarily, one or more members of the managerial staff is given the responsibility of preparing, coordinating, maintaining and disseminating both general and in-house standards.[28]

In the field of graphic presentation many organizations and agencies have prepared standard codes or compilations for internal use. These codes are frequently reproduced in the form of printed manuals which may also include specifications and standards for tabular and textual presentation. The following are a few examples: Department of the Army, Standards of Statistical Presentation, Pamphlet 325-10, April 6, 1966; National Institutes of Health, Manual of Statistical Presentation, Division of Research Grants, Statistical Item Number 10, January 1970; United States Department of Agriculture, Office of Management Services, Tips on Preparing Chart Roughs, Washington, D.C., 1973. Not infrequently, a treatise on graphic presentation may serve as an internal standards guide. When I was in charge of the Center for Studies in Demography and Ecology at the University of Washington and of the Washington State Census Board, my Handbook of Graphic Presentation along with a file of a few hundred charts from previous studies comprised our internal standards guide.[29]

The practical importance of internal or "company" standards of graphic presentation is exemplified by the following uses and benefits:

1) They serve as important guides in maintaining uniformity and consistency for many repetitive and recurring procedures.
2) They represent an important educational tool by facilitating the training and indoctrination of new employees, thereby relieving the supervisory staff of time and effort.
3) They can be helpful in enhancing the quality of work by developing procedures based on experience, collaboration and consensus.
4) They can reduce costs by increasing the efficiency and economy of basic procedures.
5) Standards represent a distillation of experience which can be retained and perpetuated without dissipating time and energy in constantly retracing or reinventing certain procedures.

Concluding Remarks

In addition to the usual conventional comments, the concluding remarks of this discussion will attempt to provide a broader perspective to the role of standards by relating standards to certain facts and issues concerning recent and future trends in graphic presentations.

In order to maintain and improve the quality of graphic presentation there is no doubt that sound and generally acceptable standards are indispensable. From a long-time perspective, I believe the chief concern of most specialists working in this area is to improve the quality, effectiveness and acceptability of graphic presentation as a graphic language--a medium of visual communication. It would be wishful thinking to assume that standards alone could achieve such a goal. Even the most theoretically acceptable, technically sound, and complete standards per se are of no consequence unless they are widely known, understood and applied. From observation over the years, one is compelled to conclude that a significantly large proportion of published charts are prepared by persons with little or no knowledge of standards, to say nothing of other aspects of graphic presentation.

Standards, though indispensable, constitute only one facet of a large body of theoretical principles, substantive facts and practical know-how which comprise the art of graphic presentation. All of these varied elements are interrelated and interdependent. In the preparation of charts, standards provide essential guides and direction, but in addition a wide variety of other knowledge as well as skills and techniques are required.

One of the most unprecedented developments in the field of graphic presentation has been the utilization of electronic computers and auxiliary equipment in the preparation of statistical charts. This development has occurred during a span of approximately two decades, with the last five years representing a period of spectacular change. Thus far, the most significant and productive application of computer techniques in graphic presentation has been statistical mapping. This fact is evidenced by a proliferation of computer mapping systems and technologies. It has been convincingly demonstrated that computerized techniques occupy a very essential place in graphic presentation, and as far as the future is concerned further applications and developments in computer technology can be expected.

As Dr. Bachi points out, a significant factor which inevitably will operate as an accelerating force is the unprecedented output of statistical information largely from the proliferation of automation.[30] It has been estimated that by the end of the next decade new information will be generated and circulated at six times the present rate and 20 to 25 times the volume of a mere fifteen years ago. In order to produce this veritable avalanche of information "the number and processing capacity of systems which are now in operation will have to be multiplied by a factor of 50 or 100."[31] As a consequence, graphic techniques should assume a more important role as a tool for interpreting and mastering large masses of data.[32]

However, it cannot be emphasized too strongly that electronic computers are no substitute for a thorough knowledge of basic theory and practice of graphic presentation. Standards or lack of standards are based on man-made decisions. No amount of sophistication in computer technology can replace personal insight, experience and expertise in the art of graphic presentation.

In recent decades, statisticians by and large have become indifferent and neglectful of graphic presentation. As far as the main stream of contemporary statistics is concerned, graphic presentation has been shunted into a marginal niche. This trend is clearly reflected in the programs of statistical societies, in statistical journals, in university courses in statistics and in treatises on statistics. For several decades the American Statistical Association had a standing committee on graphic presentation, but after 1954 and up to the present time, it has ceased to exist. As indicated previously, the American Statistical Association is no longer represented on the American National Standards Committee on Preferred Practice for the Preparation of Graphs, Charts, and Other Technical Illustrations and not a single member of this committee is affiliated with the American Statistical Association. One observes that with growing indifference and neglect, the incidence of graphic illiteracy--"graphicacy" according to Albert Biderman's neologism--among statisticians seems to have increased. This observation can be substantiated by the clumsy and amateurish charts produced by or under the direction of statisticians.[33]

However, during the past two or three years there are indications of a renascent interest in graphic presentation among many statisticians.

This trend is attested to by the work now in progress under the leadership of Albert Biderman, the newly established Council on Social Graphics, the program initiated by Roberto Bachi on Graphical Methods at the 1975 biennial meeting of the International Statistical Institute, today's program organized by Vincent Barabba as well as other activities. A climate conductive to the progressive development of graphic presentation in terms of higher standards, improved techniques, better trained specialists, and wider acceptability is more favorable today than it has been in several decades. Hopefully, the impetus that has been started will continue without serious interruption.

In concluding this discussion, I believe it would be appropriate to present a few comments concerning standards that pertain to another aspect of graphic presentation which thus far has not been mentioned. I refer specifically to professional standards--standards of professional competence. Questions relating to the progressive development of the discipline itself as well as the achievement of higher levels of professional competence are interrelated. In this connection one of the most serious impediments to the improvement of professional standards is the lack of adequate training programs in graphic presentation. Because of their crucial importance, I trust that questions relating to professional standards, training programs, as well as other needs and shortcomings of graphic presentation as a discipline will be given the attention they deserve at some future meeting.

Footnotes

*I am indebited to Stanton E. Schmid of the University of Washington for a critical reading of the manuscript and for offering constructive suggestions.

1/ Excerpted and in part reworded from International Standards Organization (ISO), and Arnold M. Rosenwald in Rowen Glie (ed.), Speaking of Standards, Boston: Cahners Publishing Company, 1972, pp. 34-35 and pp. 152-153.

2/ J. Gaillard and Madhu S. Gakhale in Rowen Glie, Ibid., p. 16.

3/ P. G. Belitos, "The Challenge of the Decimal Inch," The Magazine of Standards (April 1961), pp. 100-105.

4/ Dickson Reck (ed.), National Standards in a Modern Economy, New York: Harper and Brothers, 1956, p. ix.

5/ Jack Rogers, "Industrial Standardization, Company Programs and Practices," Highlights for the Executive, Studies in Business Policy, No. 85, National Industrial Conference Board, 1957, p. 3.

6/ Prior to August 1966, the American National Standards Institute was named the American Standards Association.

7/ William Playfair, The Commerical and Political Atlas (third edition) London: J. Wallis, 1801, pp. ix-x.

8/ Ibid., p. xii. For a more extensive discussion of Playfair, see H. I. Funkhouser and H. M. Walker, "Playfair and His Charts," Economic History, Vol. III (1935), pp. 103-109.

9/ H. I. Funkhouser, "Historical Development of the Graphical Representation of Statistical Data," Osiris, Vol. 3 (1937), p. 280.

10/ H. I. Funkhouser, Ibid., p. 271.

11/ H. I. Funkhouser, Ibid., p. 317.

12/ H. I. Funkhouser, Ibid., p. 319.

13/ Joint Committee on Standards for Graphic Presentation, "Preliminary Report Published for the Purpose of Inviting Suggestions for the Benefit of the Committee," Publications of the American Statistical Association, Vol. XIV (1914-15), pp. 790-797. This report was also published as a separate pamphlet by the American Society of Mechanical Engineers.

14/ The following summary statements were excerpted and partially re-written from American National Standards Institute procedures furnished by L. E. Farragher, Standards Engineering Administrator of the American Society of Mechanical Engineers.

15/ Committee on Standards of Graphic Presentation, Time-Series Charts: A Manual of Design and Construction, New York: The American Society of Mechanical Engineers, 1938, p. 6.

16/ Subcommittee Y15.2 of the Committee on Preferred Practice for the Preparation of Graphs, Charts and Other Technical Illustrations, American Standard Time-Series Charts, New York: American Society of Mechanical Engineers, 1960, p. 1.

17/ "At the present time, the subcommittee for Y15.2, Time Series Charts, is preparing a revision for this standard. Work is just starting on this project and little has been accomplished, as yet. The existing standard, completed in 1960, is slightly out of date and not much, in my opinion, needs to be changed."--Quoted from a letter by Francis Saint, Chairman, Y15 Committee, July 20, 1976.

18/ Executive Office of the President: Office of Management and Budget, Social Indicators, 1973, Washington, D.C.: Government Printing Office, 1973. There are 165 numbered charts in this study. Actually there are more charts since it will be found that two or three distinct charts are sometimes included under a single number.

19/ Age-and-sex pyramids, of course, are generically double histograms.

20/ Calvin F. Schmid and Stanton E. Schmid, Crime in the State of Washington, Olympia: Washington State Planning and Community Affairs Agency, 1972; Calvin F. Schmid, Charles E. Nobbe, and Arlene E. Mitchell, Nonwhite Races, State of Washington, Olympia: Washington State Planning and Community Affairs Agency, 1968; Calvin F. Schmid and Stanton E. Schmid, Growth of Towns and Cities, State of Washington, Olympia: Washington State Planning and Community Affairs Agency, 1969.

21/ Of course, in this connection it must be recognized that a certain body of standards are applicable to a greater or less degree to all graphic forms.

22/ Subcommittee Y15.2 of the Committee on Preferred Practice for the Preparation of Graphs, Charts, and Other Technical Illustrations, Loc. Cit.; Jack Rogers, Op. Cit., p. 5.

23/ Leonard P. Ayres, The War with Germany, A Statistical Summary, Washington, D.C.: Government Printing Office, 1919. There are 60 "diagrams" and 12 "maps" in the report. Judged in terms of present-day standards, most of the charts would have a high rating. However, several bar charts without scale lines and scale figures would be classified as deficient.

24/ Paul J. Fitzpatrick, "The Development of Graphic Presentation of Statistical Data in the United States," Social Science, Vol. 37 (October 1962), pp. 203-214.

25/ H. I. Funkhouser, Op. Cit., p. 375.

26/ Roxann A. Van Dusen, Social Indicators, 1973: A Review Symposium, Washington, D.C.: Social Science Research Council, 1974.

27/ Ibid., p. 68.

28/ Frequently, such functionaries are called standards engineers. Standard engineers represent an established profession with an autonomous organization and journal.

29/ Calvin F. Schmid, Handbook of Graphic Presentation, New York: The Ronald Press Company, 1954.

30/ Roberto Bachi, "Graphic Methods: Achievements and Challenges for the Future," presented at the Fortieth Session of the International Statistical Institute, Warsaw, Poland, September 1-9, 1975.

31/ Georges Anderla, "The Future of Information for Governments and Society," OECD Observer (Organization of Economic Cooperation and Development), No. 63 (April 1973), pp. 27-32.

32/ In this connection, as a significant historical note, in 1915 one of the most cogent reasons for formulating standards of graphic presentation was the increasing volume of information. According to the report of the original Joint Committee on Standards for Graphic Presentation,

> As civilization advances there is being brought to the attention of the average individual a constantly increasing volume of comparative figures and general data of a scientific, technical and statistical nature. The graphic method permits the presentation of such figures and data with a great saving of time and also with more clearness than would otherwise be obtained. If simple and convenient standards can be found and made generally known, there will be possible a more universal use of graphic methods with a consequent gain to mankind because of the greater speed and accuracy with which complex information may be imparted and interpreted. Op. Cit., pp. 1-2.

33/ Calvin F. Schmid, "Some Comments on Roberto Bachi's 'Graphical Methods: Achievements and Challenges for the Future,'" presented at the Fortieth Session of the International Statistical Institute, Warsaw, Poland, September 1-9, 1975.

AUTOMATING STATISTICAL GRAPHICS: A TOOL FOR COMMUNICATION

Vincent P. Barabba, Bureau of the Census

Today I'm going to discuss with you some efforts, both inside and outside the Census Bureau, to develop a fully automated and standardized graphic presentation system.

This is a very important endeavor. People who deal with statistics today are trying to cope with an information and data explosion that is increasing so fast that by the end of the next decade, as Cal Schmid pointed out, we may be contending with perhaps six times the present volume of statistical information. Given these realities it is imperative that we find immediately more uniform and more effective ways to communicate statistics quickly, accurately and more understandably.

What I have to say ties in directly with the presentations of my colleagues in their session. Standards and perception of graphics are inseparable from a useful, fully automated standardized system. We are all concerned with what we might call graphic literacy.

As we go through this presentation, I'm asking you to consider factors that will be useful to researchers in graphical statistics in developing a better sense of direction, not only regarding hardware and software but also in answering such questions as what statistics lend themselves to the system and how should they be presented. Although the Bureau has had little dialogue with other groups on this subject in the past, if the system is to develop effectively, we must have much dialogue in the future. We need your input because, in my judgment, the system we are developing will have a strong impact on the future use of statistics generated by the Federal statistical system.

There is at the present time little empirical evidence that graphics have a greater impact than textual or tabular material. We can say that intuitively graphics appear to be more effective; they certainly attract attention to a greater extent. Although users of data have been skeptical of the so-called "graphics method of statistics" for more than 100 years, I believe that it is time we begin, in a systematic way, to gather empirical evidence on the impact of graphics. We need to demonstrate the strengths of the various forms of graphic presentation in order to justify the investment in the system. Dr. Wainer's efforts represent a first approach to that need.

To be most useful, the systems must be standardized. Factors which may be considered as part of a standardized system include portability, adaptability to specific needs, flexibility to provide for simple as well as sophisticated usage at varying levels of complexity and financial investment, and ability to be used by people with a variety of skills and training.

What I have just said about graphics, skepticism about their use, and difficulty in arriving at suitable standards is not new. For instance, 700 years ago the Mayan Indians had a highly developed written language based almost entirely on symbols of people and faces. An interesting treatise on the development of statistical graphics is given by James R. Beniger in his unpublished paper entitled "Science's Unwritten History: The Development of Quantitative and Statistical Graphics." As Cal Schmid pointed out, the International Statistical Institute which was organized in 1885 gave serious consideration to standards of graphic presentation but with little success. That same year, the noted economist, Professor Alfred Marshall, wrote the following in the Jubilee Volume of the Statistical Society of London:

> "The graphic method of statistics, though inferior to the numerical in accuracy of representation, has the advantage of enabling the eye to take in at once a long series of facts. ...Its defects are such that many statisticians seldom use it except for the purpose of popular exposition, and for this purpose I must confess it has great dangers. I would however venture to suggest the inquiry whether the method has had a fair chance. It seems to me that so long as it is used in a desultory and unsystematic manner its faults produce their full effect, but its virtues do not."

Although many of the concerns expressed by Professor Marshall almost 100 years ago are still being echoed today, it appears that the climate in the statistical community is such that serious efforts to present data more effectively through the medium of graphics must be considered.

For many reasons, the Bureau of the Census is interested in promoting this activity. First, it is a major contributor to (some might say a culprit in) the information explosion. Second, we have a historical as well as current statistical base on which to build. Third, our users have indicated that our methods of display are cumbersome, require time to digest and have the potential for inaccurate interpretation. Fourth, the Bureau has much of the basic software and hardware that is required to develop an automated standardized system and last but far from least, we have innovative, well trained personnel to develop the procedures and demonstrate their usefulness.

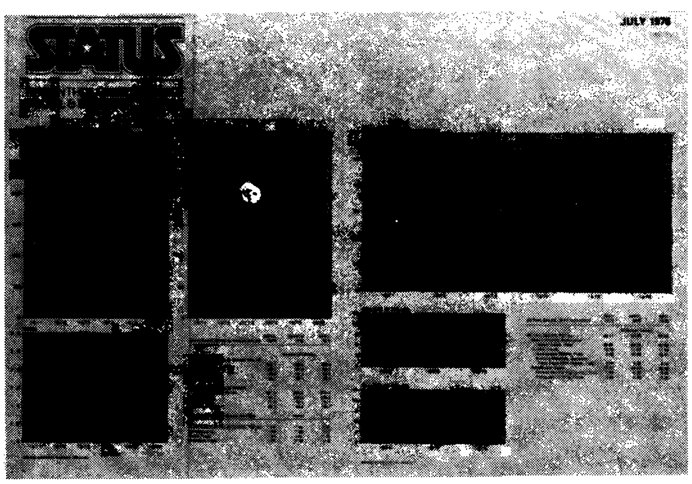
Our ultimate goal at the Bureau, which may take a few years to realize, is to put together a system that will produce statistical displays in black and white or in color, in a variety of formats. The displays include those produced electronically on a television screen and on videotape, in the form of color printouts, color slides, and microfilm, or whatever format demonstrates a potential for effective communication quickly and at reasonable cost.

We have, in fact, already made a good start in the development of this automated system for presenting statistics. One product that has resulted from our efforts thus far is a new chartbook of social and economic trends. The issue for August contains 86 pages, including both text and more than 165 charts and 4 maps -- all created by computer at the Census Bureau.

This publication grew from a series of briefing notes on domestic developments prepared by the Bureau from data compiled from the entire Federal statistical system for the President and Vice President, starting in April of last year. The graphic approach that the Bureau took impressed them to the extent that they decided that a more extensive publication, stressing the use of 4-color graphics, would be of great value not only to the Government, but also to the public.

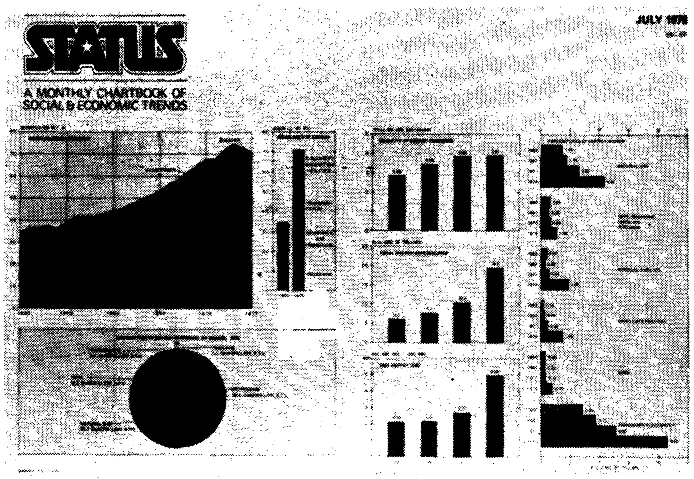
The result is our new magazine, called STATUS, which stands for Statistics, United States. On the first of July, STATUS made its formal debut in ceremonies at the Bureau commemorating the 25th birthday of UNIVAC I. Vice President Rockefeller attended, and in his remarks he pointed out that in 25 years we have progressed from the ability to simply produce more information via computer, to an important new dimension -- the actual dissemination of information in graphic form generated by the current family of computers.

Here are some line charts from STATUS. They are actually in four colors but are reproduced black and white here. It took the computerized system less than five minutes to produce each of these high-quality charts, many times less than it would take an artist to do the work. And,



of course, the more complicated the chart, the more time an artist would have to spend on them, whereas the system do can them all relatively quickly.

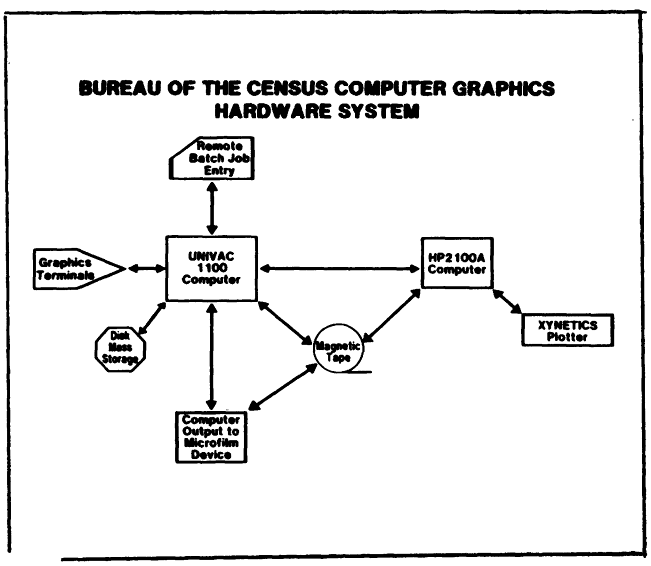
Here are some more graphics from STATUS, illustrating the system's ability to produce bar charts and pie charts as well as line charts. I might add that the comments that we already have



received on STATUS -- both in and out of Government -- have expressed a great deal of amazement at the speed at which we have been able to produce the statistics in graphic form. Data less than two weeks old, in some cases, appear graphically in published form.

Let's take a look at the computer hardware system that we use to produce these graphics.

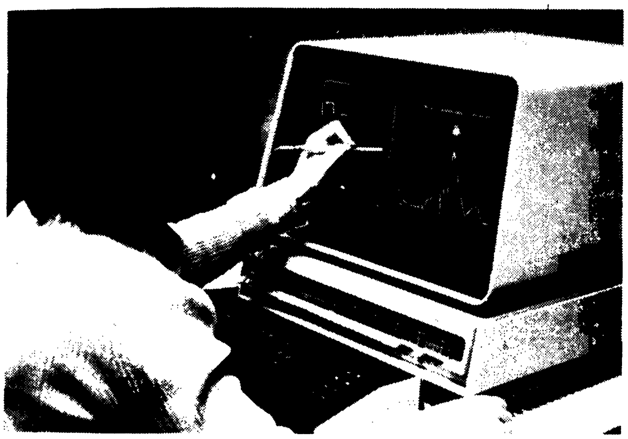
In this schematic, the boxes in the middle represent our central processors. On the left is the graphics terminal, which can feed graphical descriptions into the computer through a keyboard and display the resulting charts on a screen. The other two main pieces of hardware receive the instructions from the computer and then actually



produce the hard copy. They are the Xynetics flatbed plotter, shown on the right in the schematic, and the computer output to microfilm equipment, shown at the bottom of the schematic, which we call the COM system. The arrows simply show the data communications lines. These normally are telephone lines from a remote graphics terminal location to the computers, and from the computers to the hardware that makes the graphics.

These three main units -- the graphics terminal, the plotter and the COM system -- cost the Census Bureau about $570,000, but less sophisticated counterparts would be much less expensive. For about $50,000 you can obtain basic graphics.

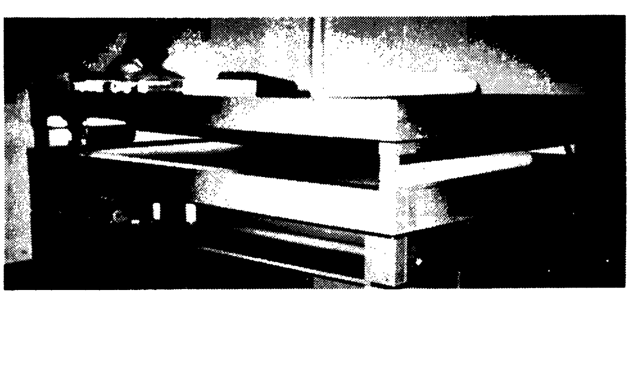
The graphics terminal consists of a cathode ray tube, or CRT as we call it, which is like a television tube, and the keyboard for issuing instructions. The operator is able to see on the screen of the CRT what is actually being cor



structed from information fed to the computer. If you don't like what you see, you can issue further instructions to the computer and change any any element in the picture.

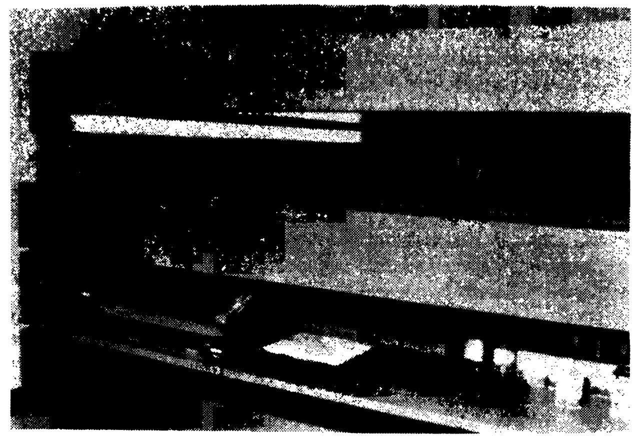
When you are satisfied, you just instruct the computer to have either of the hard copy devices produce the chart. You also can get a paper copy or copies in seconds through a copying machine that is hooked up to the CRT. This CRT runs about $10,000, and this is down from $12,000 five years ago for a terminal that had much less capability.

This is the Xynetics flatbed plotter that draws the charts from instructions received from the



computer. We acquired this 2 years ago for about $110,000. Today they run from $20,000 for a 1-pen plotter to as high as $200,000 for the most for the most sophisticated version.

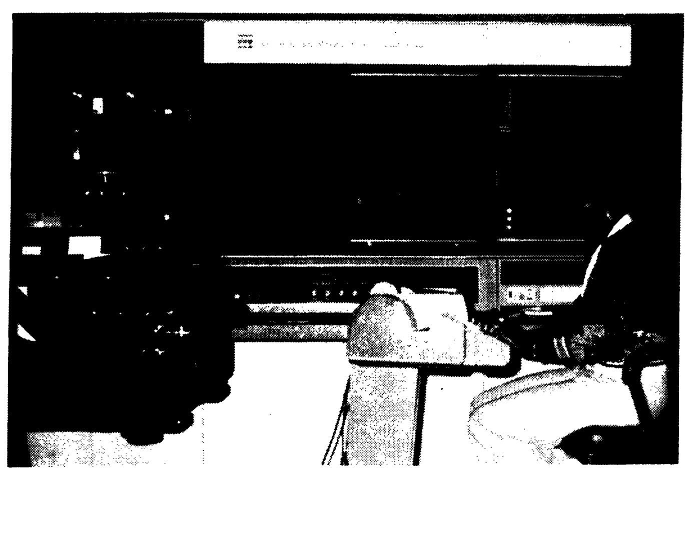
You can see here the drawing head that can produce graphics. It is program controlled and and is capable of plotting charts in different sizes,



in black, or in black plus three additional colors made by pens that can be soft tipped, ball point, or liquid ink, and that can draw on mylar, plain plain bond paper, or tracing paper.

The weekly White House briefings, which incidentally we still prepare, use charts produced by this plotter. For STATUS, however, we still have to go through color selection and separation, have the printing plates made, and then do the printing. A little later I will describe equipment that will eliminate this and directly produce full-color hard copy -- and our ability to use this new capability may not be more than a year or so away.

This is the COM system, once again the letters C-O-M standing for computer output to microfilm. On the right is the tape drive, which has a magnetic tape on it that has the instructions from the CRT via the computer. In the center is a screen that allows you to see the image of the graphic. If you like it,

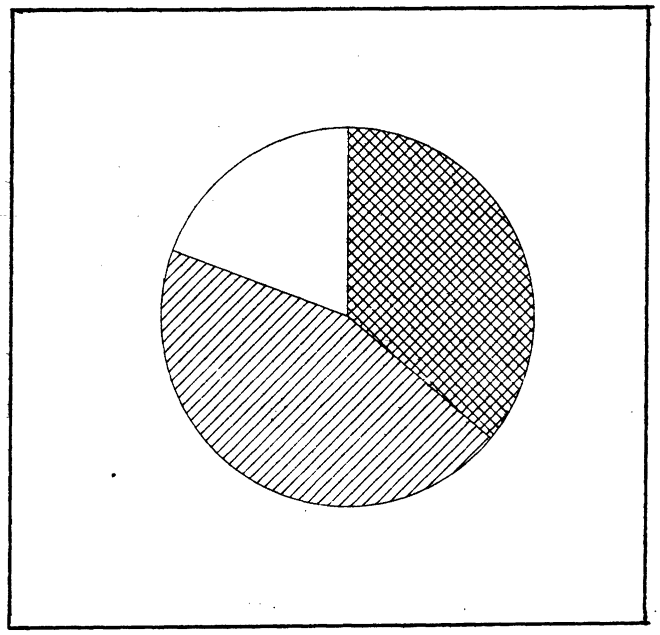


you can project the image on the screen at the left, where it is photographed by any of four cameras -- a 16 millimeter, a 35, a 105 millimeter microfiche camera, or a large 310 millimeter camera that will directly produce 8 1/2 by 11 inch pages on either photo typesetting paper, or on film.

It's important to note that the printshop still has to add colors, if they are needed, but the COM system eliminates the outside photographic process. This particular system cost $450,000, but a basic COM machine could be purchased for $150,000.

So these are the three types of graphic devices that we have now in our system. The big breakthrough is that we have added to the basic ability of the computer and have entered the custom stage where we can ask the computer to give us clear and accurate graphics made to order. Here are some examples.

For instance, here is a simple pie chart, created by the programmed computer. It has three slices with three different shadings.

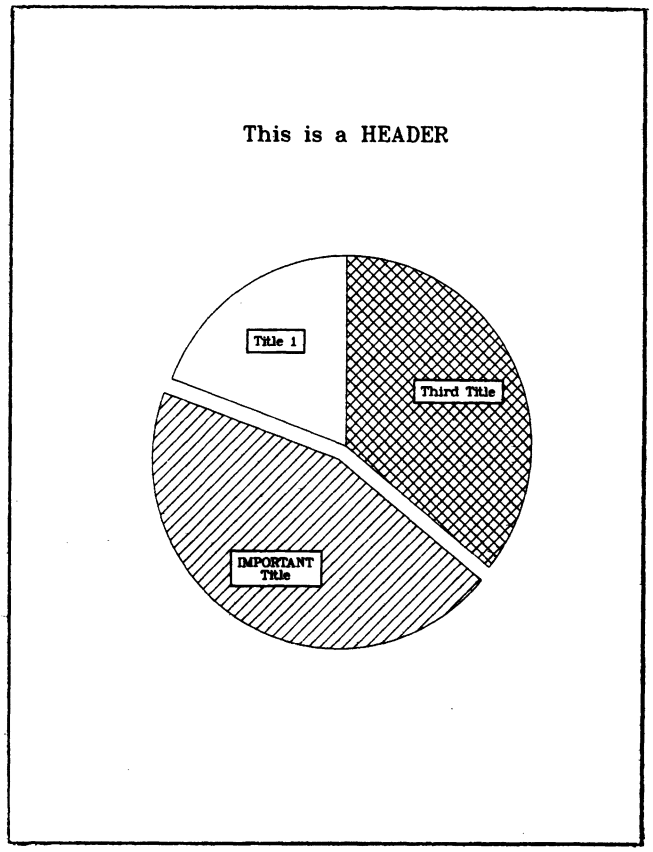


It was created through these simple, English-language instructions. The first was to lay out three slices. The word "end" means that there are no more graphic instructions. And then we gave the data values -- in this

```
# LAYOUT SLICES 3
# END
14 32.3 26
```

case 14, 32.3 and 26 -- and we generate our pie chart.

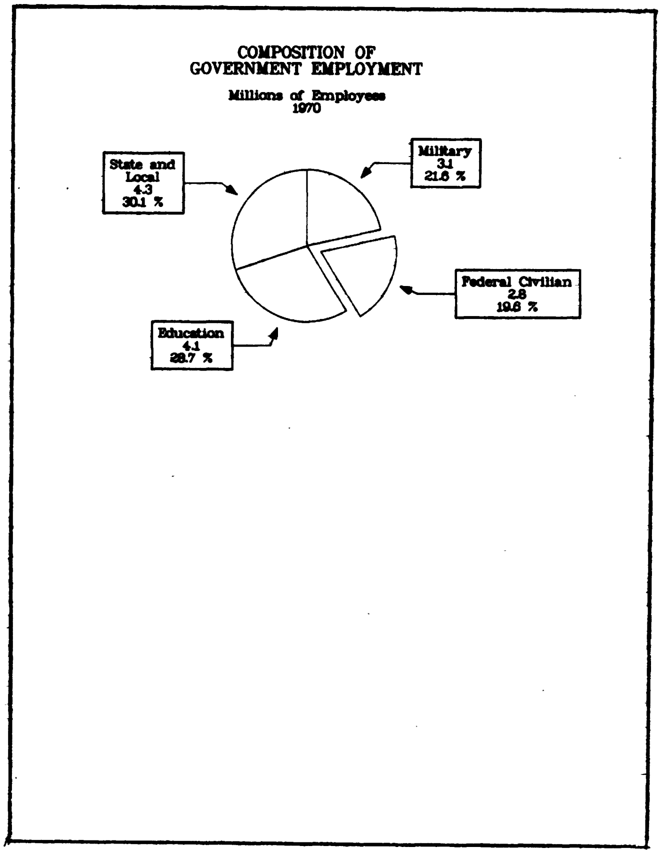
But suppose you want more sophistication. You want a separated slice; a title, or header, at the top; and titles on the slices, including one that is more important than the other two as we have here.



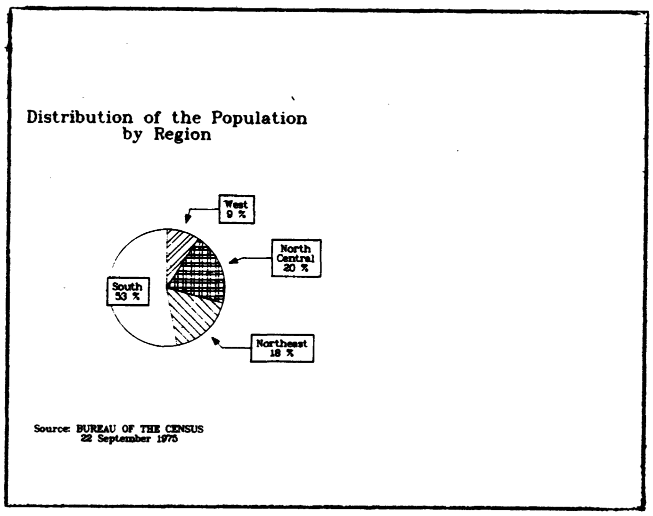
Using the same data values, these are the additional instructions that were needed to produce the chart. The computer took care of all the housekeeping. It determined the size of the pie, the size of the page, the textures, and where where to put the titles -- all through previous programming.

```
# LAYOUT SLICES 3
# BROKEN 2
# HEADING/<T>HIS IS A <HEADER>/
# SLICE TITLES/<T>ITLE 1/
      <IMPORTANT>&<T>ITLE/
  <T>HIRD <T>ITLE/
# END
14 32.3 26
```

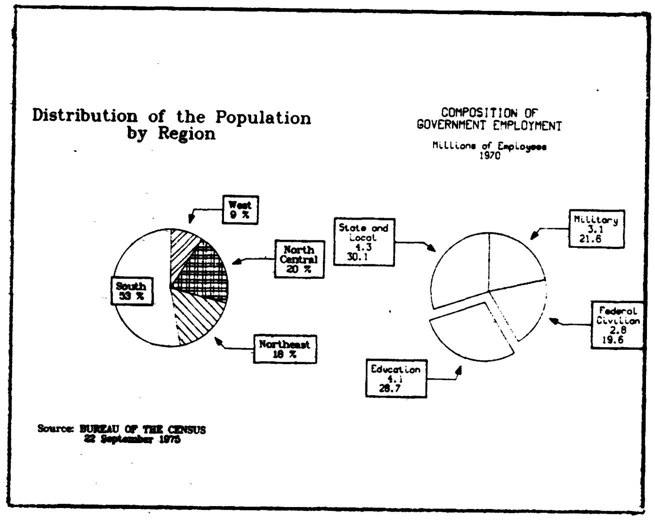
You can have the computer put the chart at the top of the page, and in this case put titles in boxes and indicate the slices with arrows.



Or you can put the chart on one side of a horizontal page. You could create text material by the computer, and strip it into an apropriate place as part of the graphic -- on the right side in this particular situation.

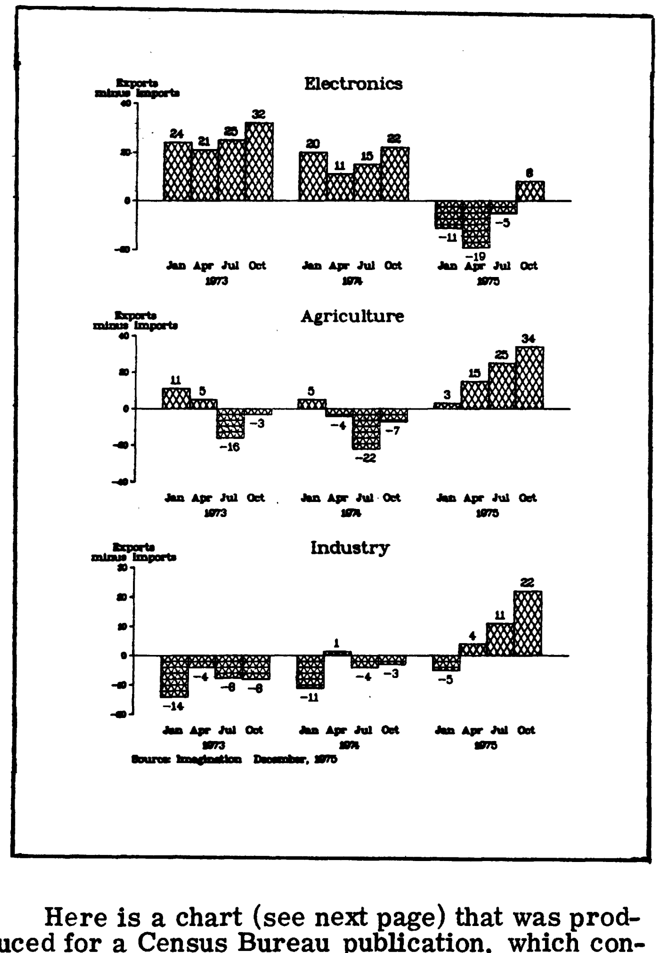


You can put charts side by side, if you want, and the computer will place all these things for you. And again, you can change what you don't like. For instance, you could have asked the com-

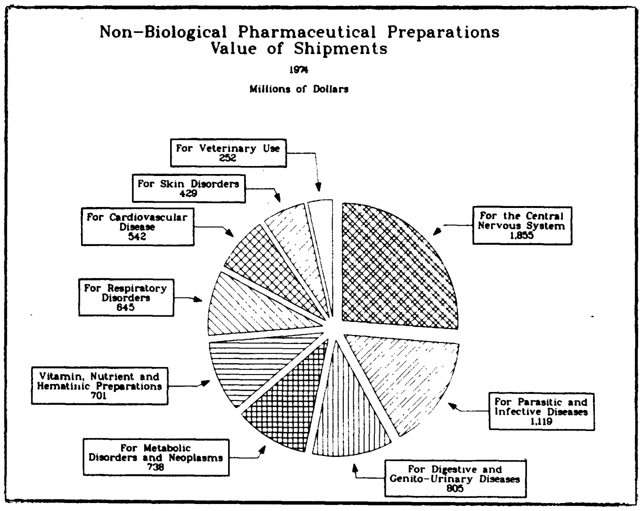


puter to move that left pie chart a little more to to the left, if you wanted to, after looking at it on the CRT. Also note here that we can put the titles in a variety of type styles.

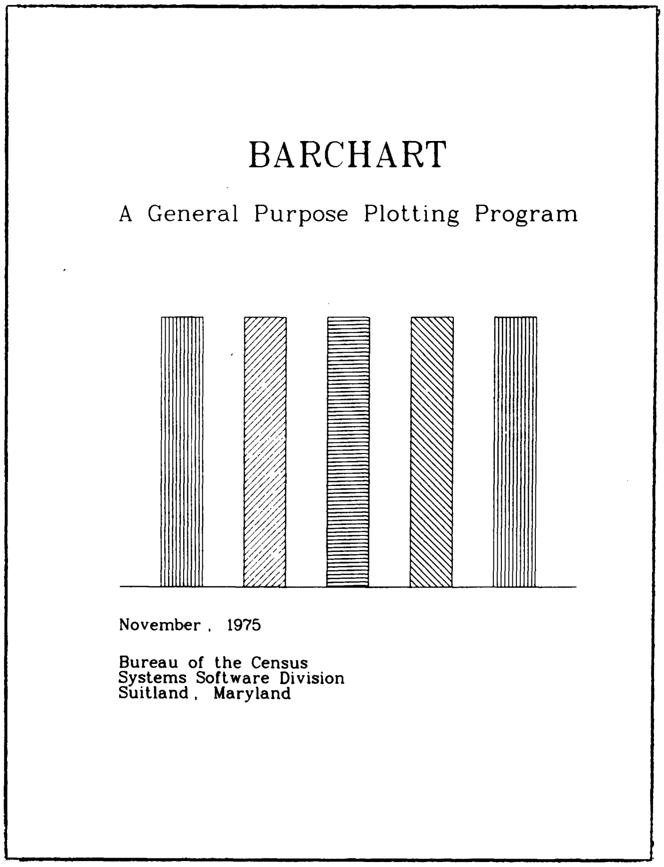
And here is a case where you have three bar charts on the same page, with scales on the left side.



Here is a chart (see next page) that was produced for a Census Bureau publication, which contains quite a bit of graphic information. You can imagine the amount of time that it would take an artist to do this particular chart, considering all the titles, boxes, arrows and hash marks.

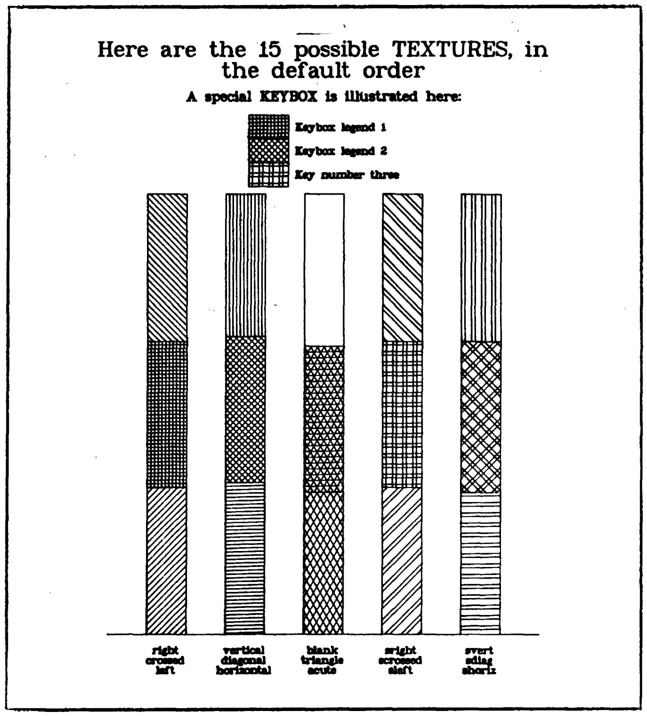


And you can get a variety of hash markings in bar charts.

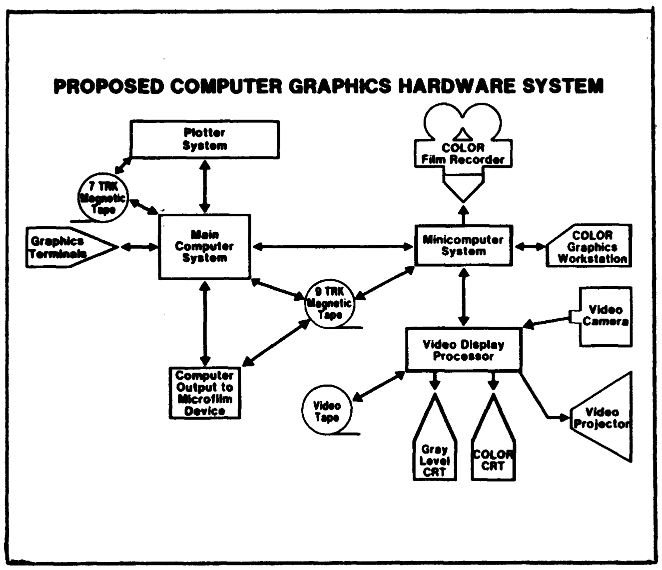


This illustration (see top of next column) shows 15 possible textures that have been programmed into our computer. Actually, an infinite variety of textures could be selected and programmed.

So this is the type of thing we can do today with the system, but we can only do it in black and white, except for the limited set of four color pens in the plotter. So what we are aiming for is to develop the ability to do these same things in full color.

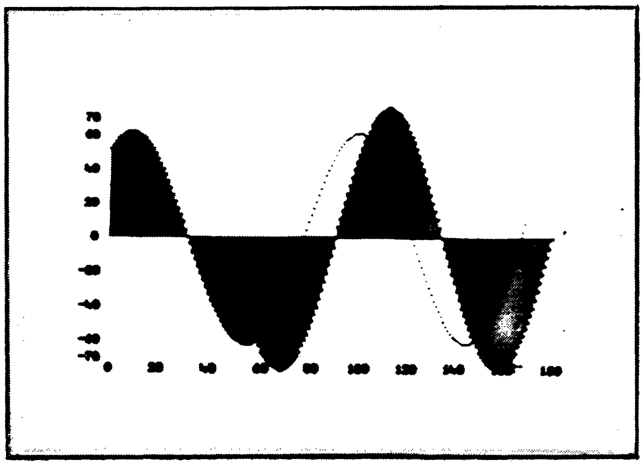


This is a schematic of our present system plus the equipment that would be needed to provide color capability. The present system is on the left, with the graphics terminal, the plotter system and the COM device. On the right is what we envision from the color standpoint. And let me point out that these color components already have been developed by companies outside the Census Bureau.



The photograph at the top of the next page shows the color CRT, although this illustration is of necessity in black and white. In addition to being able to change the composition of the graphics at will, this CRT will allow you to make color choices and quality judgments on the spot, until you are satisfied. It has, literally, a dial-a-color capability. These units would cost from $20,000 to $70,000.

Here is a chart that illustrates the display capability of a color CRT, although again it is in black and white here. This was actually produced by a color CRT developed by one of the companies. In fact this is a photograph of the CRT face. Again, the computer has done all of of the scaling and the rest of the design.



There are several devices being developed at the present time that will be able to produce hard copies in color very quickly, once you are are happy with the colors on the CRT. One of the devices would be interfaced with a color copying machine. Also, of course, you could store your approved color graphic and call it up for reproduction at any time.

Another valuable addition to a full-color system will be a video projector that will allow you to project a graphic image onto a movie-type of screen, electronically. Video equipment can can cost anywhere from $5,000 to $90,000.

Equipment also has been developed that will permit you to put images created on the color CRT screen onto color film. Equipment of this type would be in the $50,000 to $125,000 range.

One point to remember is that the graphics that will be available in color also will continue to be available in black and white. And they also will be available in tones of gray, which is not the case at present with our system.

One of the things that private industry is working on now is the development of information in a computerized data base that can tie into the rest of an automated graphics presentation system. This would permit the decision-maker to have direct interaction with the data file through the CRT or some similar approach. And this would be either in color or in black and white or gray, and a hard copy of what he wants would be produced by issuing a simple instruction.

So these are some of the things that we have today, either at the Census Bureau or in the research laboratories of private industry. As is illustrated by STATUS magazine, we already have the automated tools to produce effective and accurate graphics, and we hope within a year to have acquired the equipment to create color as well as black and white, both electronically and in hard copies of different kinds. Then we will have to put the entire system together, which should take a little more time.

Ultimately, we would like to see this system, or one of a modified nature, made available to users in any location, at reasonable cost and in a standardized form that will permit a maximum amount of flexibility. In the meantime, we will encourage and participate in research to determine determine the extent of the impact that graphics have on people, and if necessary take account of this impact as we continue to develop the system.

ASSESSING THE EFFICACY OF VISUAL DISPLAYS

Howard Wainer and Mark Reiser, The University of Chicago

Abstract

In an effort to measure the efficacy of several types of graphical displays, an experiment was performed in which a question was asked of the subject and his/her response time was measured. It was felt that any unambiguous display would allow the correct response eventually, and so response time seemed a natural dependent variable. The results indicated that for the type of question asked there was a definite order of preference with a tabular representation finishing last. This experiment is used as an exemplar of the difficulties involved in the empirical study of this problem area.

Introduction

On the surface the question, "Which of these two displays is better?" appears to be a perfectly reasonable one. Moreover, it initially appears that one could answer it using the traditional scaling methods that psychometricians have been using since the time of Thurstone. Sadly, a unique answer to this question is as difficult to determine as the answer to the question, "Which of two estimators is better?" In both cases there is no unique answer; it depends upon the situation. More precisely, one must specify the question in more detail before an answer can be found. Thus we must not ask which of two displays is better, but rather which is better for yielding the answer to a particular question. When we have reached the point where we know precisely what information we want to represent in a display we can easily test various candidates as to the clarity with which they convey that information. Previously we studied (Wainer, 1974) the efficacy of hanging histograms which have been proposed by John Tukey (Tukey, 1977) as an improvement on standard histograms. In that study we used a variant of Fechner's method of paired comparisons which is generally called (Bock & Jones, 1968) the constant method. In this task a series of stimuli are paired, one at a time, with a constant one, and the subjects are required to judge which member of the pair is x'er, where x is the underlying dimension of interest. The various stimuli are then arranged in the order of the general frequency in which they were preferred: the one which was preferred to the constant most frequently is considered the highest, and the one over which the constant was preferred most frequently is the lowest. The proportion of times a particular stimuli is preferred is converted to a scale value through an inverse normal transformation and the scale values are (through the intervention of a Thurstonian scaling model) considered to be intervally scaled. In the particular problem for which this technique was employed, the scale value of any of the various displays would reflect the extent to which that display emphasized the dimension x on which the subjects were judging. If that dimension was utterly lost in the display, then a plot of the physical values of the dimension against the subjective scale values would have a zero slope. If it was perfectly displayed the slope would be very steep indeed. Thus one could compare two (or more) display types on the basis of their slopes. Moreover, one could also parameterize each display type by its slope <u>for that dimension</u>. Thus a future data displayer could determine what display would best suit his needs by searching through a list which would detail the specifications of each display.

Two shortcomings of such an investigative method is that the experiment is rather tedious, and one may get slopes which are not uniform across the entire dimension. Furthermore, some questions which one might ask of a display do not lend themselves to this sort of scheme.

Another method for evaluating the efficacy of a display would be to determine the amount of time that it takes a person to extract a particular bit of information from the display. This seems like a reasonable way to go about this problem since any display should, at the very least, be unambiguous, allowing the average reader to extract the information contained in it. It seems reasonable to maintain that a better display will allow faster extraction.

Some months ago we had the good fortune to participate in a workshop on display techniques during which we looked carefully at the displays used in <u>Social Indicators: 1973</u>. After much complaining about some of the displays we tried to come up with alternatives which would be better. One display which suffered grievously at our hands was Chart 2/9 which is generally called a "bar chart," and which I leave to you to judge on the

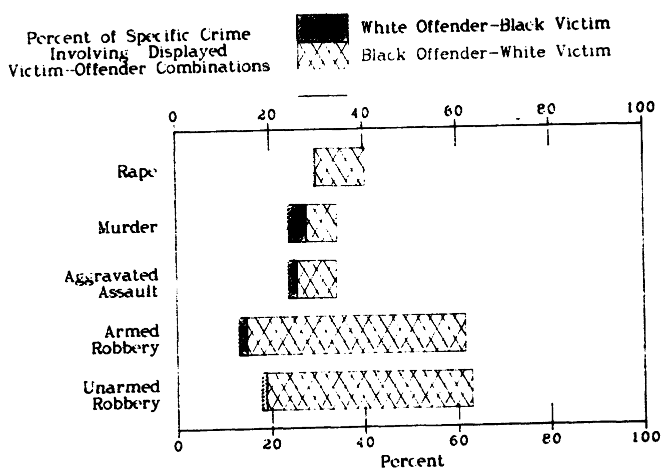


FIGURE 1. Bar chart showing race of victim and offender, by type of violent crime: 1967 (for 17 major cities).

face of it whether it deserved our derision. Stephen Fienberg conceived of one alternative which he calls a "Floating Four-Fold Circular Display" (FCD), a sample of which is displayed in Figure 2. I contributed my own version of 2/9

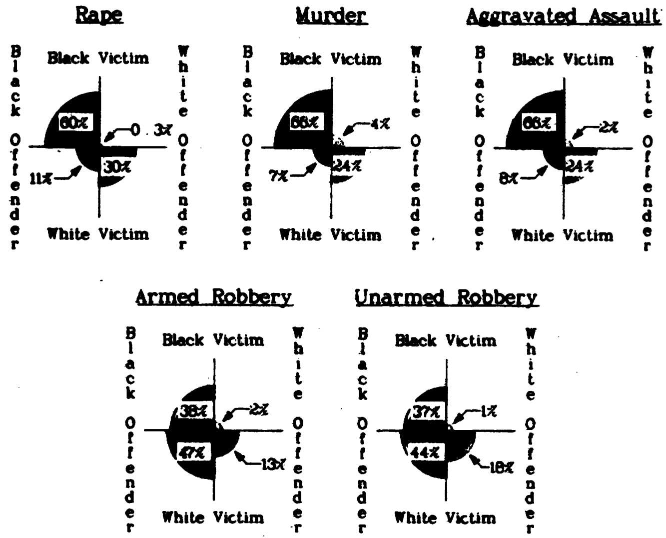


FIGURE 2. Floating four-fold contingency display showing race of victim and offender, by type of violent crime: 1967 (for 17 major cities).

which, for lack of a better name, I have denoted "cartesian rectangles." All three displays contain the same information, but each emphasizes a

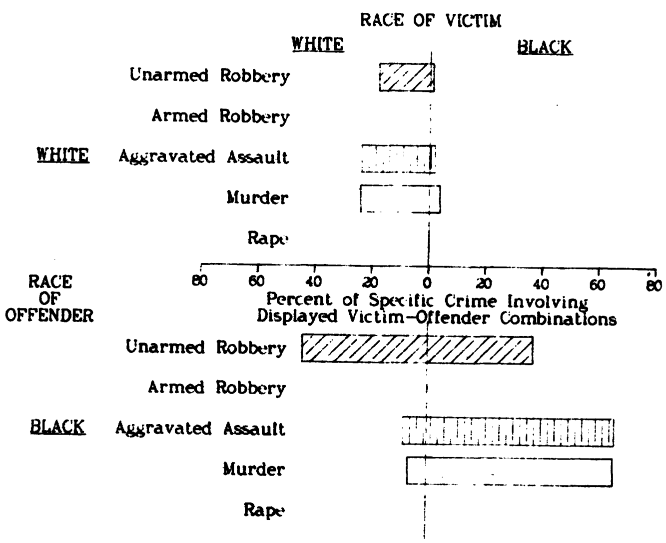


FIGURE 3. Cartesian rectangle display showing race of victim and offender, by type of violent crime: 1967 (for 17 major cities).

particular aspect of it somewhat differently. Clearly, to thoroughly test the efficacy of all these displays one would have to try a variety of different questions. Resource limitations prevented this, so the experiment which we shall describe shortly is limited in its implications, yet the methodology is generalizable. In addition to these three displays, we used as a standard of comparison a tabular display of the same data. It was felt that any saltworthy display should do better than a table of numbers.

(Percent)	White Victim		Black Victim	
Crime	White Offender	Black Offender	White Offender	Black Offender
Murder and nonnegligent manslaughter	24.0	6.5	3 8	65.7
Aggravated Assault	23 9	8.4	1.8	65.9
Forcible Rape	29 6	10.5	0.3	59 6
Armed Robbery	13 2	46.7	1.7	38.4
Unarmed Robbery	17.9	43 9	1.1	37.1

TABLE I. Race of victim and offender, by type of violent crime: 1967 (for 17 major cities).

The experiment

Sixteen right-handed students at the University of Chicago participated in the first phase of the experiment. An assertive statement was presented to the subject, and then followed by one of the displays. Each statement took the form: "In the crime of armed robbery (rape, aggravated assault), white (black) criminals victimize whites (blacks) more often than they victimize blacks (whites)." The subject's task was to decide whether the statement was true or false, based on the information in the display. Subjects were to indicate their response by pressing one button for true and another one for false.

Before the experiment, subjects were told that some of the displays represented fictitious data, so they would not be able to respond based on any previous knowledge. In fact, two sets of data were used, one real and one fictitious. Two displays of each type were made, one from each set of data. On the odd numbered trials the subjects were presented with a statement and then a display from the veritical data set. On the even numbered trials, subjects were presented with the same statement and display type as in the preceding odd numbered trial, but the display portrayed the fictitious data instead. Thus, the odd numbered trials could be considered practice or training trials, although the subjects were told that there were no practice trials. Four statements were used; thus eight trials were required for each subject. Across subjects, each question was paired with each display, and question-display pairs were balanced for order of presentation. All statements and displays were presented tachistoscopically, and responses were timed electronically. The dependent variable was response time, although response speed (1/time) was also calculated. In addition, when the presentation trials were completed, each subject was asked to order the displays from the one which he thought was easiest to use to the one that he thought was hardest.

The results

Table II represents the results of the experiment. Note that the means yield a very peculiar artifact in that on the second trial the

TABLE II

Summary of Results

	Untransformed Data		
	Mean	Midmean	Standard Error
First Trial			
Rectangles	24.46	23.12	3.55
Bar chart	20.96	18.79	2.42
FCD	31.82	25.40	5.97
Table	28.74	22.20	5.99
Second Trial			
Rectangles	11.28	10.51	1.08
Bar chart	11.59	11.01	1.53
FCD	17.37	14.35	3.80
Table	16.55	16.61	1.61
	Neg. Inverse Transformed		
First Trial			
Rectangles	-.0512	-.0444	.006
Bar chart	-.0553	-.0537	.005
FCD	-.0450	-.0409	.006
Table	-.0491	-.0460	.006
Second Trial			
Rectangles	-.1025	-.0971	.01
Bar chart	-.1093	-.0915	.015
FCD	-.1084	-.0732	.022
Table	-.0696	-.0612	.007
	Judged Preference (S.E.)		
Rectangles	2.1 (.95)		
Bar chart	2.2 (.83)		
FCD	3.1 (1.1)		
Table	2.5 (1.1)		

mean response time for the table is 16.55 seconds and for the Four-fold Contingency Display 17.37, thus indicating that the FCD is worse than the table of numbers. However, after an inverse transformation to speed we see that the table's mean speed is .0696 sec^{-1} and the FCD's is .1093 sec^{-1}. A reversal! This is caused by some outliers in the FCD. A truer picture of what is happening is seen in the mid-means (25% trimmed means) which still shows the cartesian rectangles as the display of choice (for this kind of question) followed closely by the bar charts. Next we have the FCD, while the table of numbers brings up the rear. Although the order of displays for the second trial is not the same as in the first trial, the display occasion effect was not significant as can be seen in Table III. It seems that the data from the second trial should be the more reasonable one from which to draw conclusions, since it represents subjects' performance after some practice.

The judged preferences in Table II were obtained by averaging the subjects' orderings of the graphs, where 1 indicated that the subject thought that the graph was easiest to use and 4 indicated that the subject thought the graph was hardest to use. The subjects' preferences seem to agree, in spirit, with the results from the analysis of time. The FCD was judged hardest to use, but this judgment may reflect unfamiliarity more than difficulty of use.

TABLE III

Analysis of Variance

Analysis of Variance for Reaction Time

Source	Sum of Squares	DF	Probability
Grand Mean	52992	1	.0001
Display	1472	3	.0255
Occasion	4839	1	.0001
Subjects	10393	15	.0001
Display X Occasion	112	3	.8641
Error	15957	105	
Total	85767	128	

Analysis of Variance for Speed

Source	Sum of Squares	DF	Probability
Grand Mean	.6954	1	.0001
Display	.0096	3	.1179
Occasion	.0720	1	.0001
Subjects	.0780	15	.0002
Display X Occasion	.0084	3	.1597
Error	.17	105	
Total	1.03	128	

Table III gives a summary of the analysis. Although the effect of display is significant when reaction time is the dependent variable, unfortunately, it is not significant when the dependent variable is transformed to speed. The change is due, no doubt, to large outliers in the data from the first trials. The most likely remedy is to obtain more data in order to yield more stable results, i.e., smaller error variance.

Discussion

The type of question was formulated to give an advantage to the bar chart, since a more quantitative question (e.g., "True or False--16.5% of all rapes are Whites on Blacks.") would be far easier to answer with the other displays wherein these numbers can be read directly rather than obtained through subtraction of two points on the x-axis. For this type of question the table might do better still.

Even with this edge the bar chart did not run away from the competition. We thus conclude that even for this situation (that is, the one in which it seems best suited) the bar chart is not the easiest to use, and other display types are to be preferred. It is interesting to note further that the bar chart was apparently the most familiar display to the subjects since in the training trial they responded most rapidly to it. Even with this advantage it still did not win. We feel that the more innovative display types would do even better with a more extensive training period.

It seems to us that a catalogue of display types could be prepared (much as Cal Schmid has done in his Handbook of Graphic Presentation) which would not only include categorizations of various displays but also some sort of parameterization indicating how good each display type is for each of a variety of purposes. The prospective user could then reach into this bag and pull out the one which most nearly fills all of his needs. This has two interesting sidelights. First, it implies that a great deal of empirical work of the sort we have tried to illustrate must be done, although one would hope that it would be done with greater experimental cleverness than we were able to muster. Second, it places an additional load on the prospective displayer to explicitly determine what particular aspect of his data he is most interested in emphasizing. The displayer's emphasis can be determined by his readers by checking back to the gedanken handbook mentioned previously to see what aspect of the data the display of choice was supposed to emphasize.

We believe that investigations using such methods as multi-dimentional scaling could be very useful in determining the perceptual dimensions that are involved in the perception of a display. This would facilitate the development of a theoretical structure of display construction. In addition, it may be that different kinds of audiences respond differentially to different kinds of displays; thus the individual differences models for scaling (both uni- and multi-dimensional sorts may be very helpful.

We are pursuing this currently, but we're insufficiently far along to be able to report on it at this meeting. Perhaps later. We should not end without a short comment on general graphics with no particular purpose in mind. A picture is probably the best way of finding something for which one is not explicitly looking, and so it may be that there are some general purpose display techniques which are not as good as a special purpose display for a particular question but yet are all around good performers--a graphical equivalent to the Jackknife. Whereas the all around good performers should be treasured for exploration, special purpose displays seem to be required for investigation in depth. Just as the Jackknife can drive screws (or test regression coefficients) a special tool can do each of these special jobs better. Thus, if a special job is at hand one should use the right tool, but if there is a family of jobs, one should use the most efficacious general tool. Just which is which can only be obtained through careful empirical work.

References

Bock, R. D., & Jones, L. V. The measurement and prediction of judgment and choice. San Francisco: Holden-Day, 1968.

Schmid, C. F. Handbook of Graphic Presentation. New York: The Ronald Press Company, 1954.

Tukey, J. W. Exploratory data analysis. Reading, Mass.: Addison-Wesley, 1977.

Wainer, H. The suspended rootogram and other visual displays: An empirical validation. The American Statistician, 1974, 28, 143-145.

DISCUSSION

Albert D. Biderman, Bureau of Social Science Research, Inc.

This session marks a rebirth of serious and prominent attention in the ASA to graphics for the analysis and communication of statistics. It is extremely exciting to participate in a session organized by a Director of the Census who is reviving the tradition of his illustrious predecessors Francis Walker and Henry Gannett and their great graphic presentations of the 9th through 12th Censuses. The Census Bureau for the first time since the 12th Census is again at the forefront of statistical graphic innovation. I am similarly honored at sharing the platform with Roberto Bachi who is responsible for the revival last year of the great statistical graphic tradition of the International Statistical Institute after an unfortunately prolonged period of dormancy within that body. To Cal Schmid, I am particularly indebted for his having been the major bearer of the graphic tradition within my own field, sociology, during my entire professional life. And, not the least noteworthy representation on this panel is that of Howard Wainer, for the integral place graphics now has in exploratory data analysis as well as for his work on the essential links that will have to be sustained between statistical graphics and cognitive and perceptual psychology.

Three related tensions that have long existed in graphics are raised in this session. The first is that between establishing and adhering to a restricted set of conventional standards of graphic representation as opposed to the desire to bring the full range of one's ingenuity to bear anew on each problem's unique demands. Second, is the tension between graphic simplicity and complex reality. Finally, there is the tension in seeking rules for graphic practice between relying on deduction or intuition versus explicit testing.

Standardization vs. Innovation

First, there is a need to restrain the innovative and creative impulse, even where some altogether new departure appears clearly an improvement over standard practice. Observing conventions is valuable even if their neglect poses no peril of wasteful "reinvention of the wheel" or of one's being oblivious of important considerations that are built into those graphic forms that have selectively survived to become standard practice. At the same time there is indeed vast need and room for innovative creation in social graphics. Furthermore, such standards as we have, of any of the forms mentioned in Schmid's paper, are silent with regard to many critical points of decision regarding how to proceed with the graphic treatment of any set of data for any particular purpose. Those we do have also rest almost exclusively on deductive principles, because, despite spurts of experimental activity, the sum total of experimentally-based knowledge is small and, as often as not, misleading.

Simplification vs. Oversimplicity

The second tension arises from the social scientist's desire to use graphics to convey the full complexity of his subject matter--a desire that we hope is matched by a resistance among his audience to patent oversimplifications. In social science, the importance attached to telling (and being told) the whole, complex and frequently messy story conflicts with the fact that the virtues of graphs usually vary directly with their neatness and simplicity.

Vincent Barabba and his colleagues at the Census Bureau, are exploring possibilities that the state of the art now opens up for coping with what previously have been hopelessly complex and intractable problems. Realizing these possibilities will benefit from graphic innovation--new graphics adapted to new forms of statistical analysis, as well as the mutual adaptation of modes of graphic representation and new computer-graphic reproduction technologies. The very design and production of statistical series undoubtedly have been restricted by implicit acceptance of the limitations of the media available to display the final product. If there can be radical improvements in how social statistical knowledge can be presented, perhaps they can be matched by increased sophistication in the knowledge reports attempt to present. In addition, those engaged in the production of statistical series may find new and greater degrees of order in the facts with which they deal, as they come to work with displays of these data that mobilize more of the potentialities of modern communication modalities.

Wainer has elsewhere contributed to this objective by both deductive design and experimental testing. His present experiment, however, is trivial in ways in addition to those he has discussed in his paper. As a criterion, he chose: "the amount of time that it takes a person to extract a particular bit of information from the display," that is, to reach the correct decision that a complex declarative sentence was true or false (his sentences are "complex" in the grammatical sense, although fairly simple in the sense of cognitive difficulty). Such a test, first of all, would be a useful one for graphics that have the purpose of a reference table--a convenient device, such as a railroad time table, for locating a given datum or subset in a larger data collection. (Actually, the subjects in Wainer's experiment were given an hypothesis to look-up in the display to determine whether the display agreed with it.) The graphic representations Wainer sought to test for their "goodness" are indeed ones designed for a book of "general purpose" statistics that its principal author found does have some such reference table uses--for example, by political speech-writers. But the major intent of "social indicators" chartbooks is different: it is to present data which hold within them a pattern of many patterns, that can be appreciated and learned as wholes. They are there to figure in the active work of the mind of the user, mobilizable for stimulation of and assimilation to other patterned information in that great storage and retrieval system, the nervous system. While it would not be accurate to say that Wainer asks his S's to see the trees, rather than the forest, he does ask them to report only the relative sizes of clumps and not the forest's patterns. I would like to see tests of the richness and veridicality (or even plausibility) of the stories about important social realities that subjects are able to construct while studying various sets of displays, or after they have studied them. Such tests would be closer the purposes which have led to the investments in equipment and software being made by the Census Bureau to aid the work of capturing ever more complex features of social reality in graphical displays.

Finally, and crucially, if we are to use simple prose as the criterion against which to judge the efficacy of various displays, should we not include prose as one of the display forms against which we test others?

It may be important for this matter of detail versus pattern to make explicit what I think is implicit in Wainer's rationale for making rapidity of recognition of particular relationships his criterion. Without being at all familiar with psychological research on the matter, I find it plausible that the ability to discern larger patterns is a function of the ease (which perhaps equals rapidity) with which their components can be discerned. We know from experience, for example, that if it takes too long (involves too much effort) to learn the legend of a chart thoroughly, the patterns in that chart will not be understood. But how much time or effort is too much and what range of time makes any important difference for pattern comprehension? Where on such scales do the mean times Wainer found (about 16 seconds) fall?

There is a feature of graphics that differentiates it from the verbal language and makes us more dependent on experiments in order to estimate what displays will be unduly complex for some or all of a given audience. We have poor sense of people's graphicacy because graphic communication does not afford the feedback we get with verbal communication. Indeed, most communication about graphics, as in the case with the present discussion, proceeds verbally, not graphically.

People are able to make judgments about others' literacy from listening to talk--that is, from others' linguistic fluency. They are not similarly exposed to graphic productions of others such as would allow them to make judgments of others' graphicacy. One also can form judgments, although with some degree of error, of the command of the language others possess by their reactions to what one says--signs of comprehending, not comprehending, miscomprehending--in conversations, From an accumulation of impressions of others' fluency, one forms judgments as to the general distribution of ideas and for which classes of people.

Fluency is an imperfect test of literacy; the ability to gain comprehension from prose may be greater than the ability to articulate that comprehension. Fluency is an even more imperfect measure of graphicacy, particularly so because of the limits of language about graphic form. Tests may confuse the vocabulary at a person's command for naming properties of graphics with his facility and accuracy of understanding. Good tests of graphicacy might involve asking subjects to draw patterns rather than to check words. Since the ability to draw a form that one can mentally comprehend may be limited, however, tests might best be by multiple choice among presented figures. Interactive computergraphics holds forth promise of graphic conversations that may be illuminating with regard to graphicacy.

One's ideas about what others will see in a chart, what will be clear or obscure, be attractive or not, have this meaning or that--derive mostly from projection. We think others will see as we see, although possibly not as well. Some of us may choose to employ a "test idiot" to try out a chart--"If my secretary (wife, office partner, janitor, etc.) can understand it, anybody can." Prof. Bachi used his grandchildren, possibly a large group of subjects but a suspect one if there is even the tiniest bit of genetic or intrafamilial cultural transmission of graphicacy. There is considerable warrant for the belief that all human minds are similarly constructed, and also for that fact that those likely to be an audience for a particular chart have had their minds trained in similar ways by a body of common cultural experiences. There is also warrant for the idea that some graphic forms are more pictorially literal and hence "specific-culture free" than is the case with linguistic productions. After all, we can "dig" ancient Chinese graphics and even prehistoric cavemen's drawings, but we would be lost in the corresponding verbal languages. But, at the same time, the fit of graphics to minds and cultures also varies widely. We are very apt to be surprised by what others stubbornly misperceive in a chart as by what gets perceived as intended. How are we to know what to expect of all people or of some particular set of people? What forms depend only on universal or near-universal innate ability for their comprehension; which ones on merely the knowledge expected of any reasonably educated person in modern civilization; which ones on specific training?

One of the truly remarkable lessons of the history of statistical graphics is the leap in the inventions of just one man--William Playfair--from a situation in which there was, to all intents and purposes, no such thing as a statistical graphics to where a major portion of all the basic forms most commonly used at present came into being and with remarkable similarity to current usage. They received immediate, wide acclaim. Should we page Chomsky?

Reviewing the field of experimentation on the comprehension of statistical graphics, a fairly fashionable line of endeavor a few decades back, leads me to suspect low promise from experiments like Wainer's because little generalization is possible with any confidence from the specific subjects, displays, and criteria to the larger classes each purports to represent. Most past experiments have hardly more generalizability than does advertisement copy testing. The major pertinent exceptions in our field are low-order psychophysical kinds of inquiries on grey scales, line thickness, and color valuing, as done particularly by cartographers. (See, Feinberg and Franklin, 1975.)

The graphs Wainer used, for example, all embody much more information and much higher levels of precision than required for answering the questions he put to S's. They contained high redundancy, for example, graphic and alphanumeric representations of the same quantities. With regard to each form he tested, questions might be raised whether these particular graphs are either ideal or representative of all possible designs of that general named form for conveying the data; for example, positioning and typefaces used for labels, color choices (singly and combinations), use of rulings, scale proportions. Next, we have to ask whether the particular uses, for which that form is ideal. In a forthcoming paper, Macdonald-Ross of the British Open University cogently criticizes past experiments on the effectiveness of graphs and tables on these grounds.

To even begin to replace deduction with induction as a footing for the basic graphic repertoire, and for systematic and effective testing of the rapid innovation of new and complex graphic forms now taking place, it would be hopeless to attempt to proceed in a naive empirical fashion. For experimentation to contribute would appear to require relating psychological and psychophysical theories to a more generalized conceptualization of the problems of graphic communication. Before the standards handbooks Schmid recommends would profit much from such research there would remain considerable bridge-building to relate such general knowledge to the concrete problems of graphic design confronted in any given case.

Standards and Conventions

Although innovation can be valuable, and much innovation will be essential, I agree with the arguments implicit in Schmid's paper for caution in attempting innovation of graphic forms. Funkhouser describes vast labors toward invention of new devices and elaborate systems that were in vain. Some work has been idle in that it failed to take account of the versatility of the basic standard repertoire of representational forms, and, where innovations represented any improvement at all, these were often quite marginal relative to the new learning the innovation demanded of basically lazy users and audiences. I am sure Prof. Bachi has encountered barriers to the acceptance of his Graphic Rational Patterns, even though the symbols of his system can be learned in minutes.

An equally great hazard in innovations stems from the ready possibilities for communication error and failure in graphics. An advantage of the highly familiar forms has been the accumulation of experience in how many things can go wrong with them, and, to the extent that such experience has been integrated into formal rules of standard statistical graphical representation, it figures in general caveats in manuals and in the working knowledge of regular users.

The same two types of problems beset attempts at technological innovation and the adoption of new technology for generating and transmitting graphic statistics. The problems are aggravated in that specialists who are at work in such fields as the development of computergraphics hardware and software are rarely intimately conversant with the field of statistical graphics. Not only are wheels painfully reinvented, they often turn out to be somewhat square. This fact adds urgency to the need for establishing working relationships of technological innovators with those statisticians and other scientists who have made statistical graphics a special object of their attention. This has been as objective of our Graphic Social Reporting Project (graphs) at the Bureau of Social Science Research and of the Council on Social Graphics.

There is another reason for cautions about innovation. For graphicacy to develop, both among producers and audiences, there must be the cultivation of familiarity with a restricted, durable broadly applicable set of graphic conventions and standards for social statistics. The same basic forms must be used regularly to establish them as readily drawn-upon repertoires. In the repertoire of the producer, there must be thorough appreciation of statistical-graphic forms, and the ready, economical means for producing them that frequent use generates. In the repertoire of consumers, there must be the ready and accurate comprehension that comes with the familiarity Wainer suspects affected his results.

Staying within the bounds of conventions ordinarily need not prove highly constraining. Even the most familiar basic forms are adaptable to the expression of a rich array of meanings.

Scope of Innovation

Until there is widespread familiarity with the basic statistical-graphic equivalents of orthography, vocabulary and syntax of the verbal language, innovation of new graphic forms might well be restricted to enabling one to do things that would otherwise be nigh infeasible. As a test to be applied within our project to any proposed departure from standard forms of graphical representation, I have recommended the following principle: "Does the innovation have such strong virtues and broad applicability that it has some real chance of becoming a standard--a graphic convention?" The innovative potential of the new development may derive from its elegance as a new semiotic tool of principle or its adaptability to new developments in the environments of graphic technology and uses. Exceptions to the caution against unique forms admittedly exist where an ad hoc graphic solution elegantly mobilizes for its problem the properties of "self-evidence" that some iconic devices can have. Exceptions may also exist where the statistical solution must be unconventional. I think the Census Bureau's work is embodying this strategy. The broad audience for its publications and its central role as statistical institution gives it the potential of rapidly moving innovations such as Bachi's to the status of conventions.

For all the caution in my discussion and in each of the other paper's, including Barabba's, I trust the latter's audio-visual show convinced you that the statistical graphics revolution is already here. We may not all be able to rush out to place orders for many items in the Sear's catalog Barabba put on the screen, but just a few of these systems alone have voracious appetites for the ideas we can generate to feed them and will have prodigious outputs requiring appraisal.

Here's an anecdote revealing with regard to the nature of the graphic revolution. I receive the Federal Statistics Users Conference Newsletter which, as one of its attractive features, gives me very prompt information each month on the availability of new Federal statistical publications. In the most recent issue, it contained information on new releases of school enrollment and educational attainment studies. But four days before I received my newsletter, I had already had access to these data in the multicolored chart formats of STATUS. It won't be long before they will also be available on my CRT by pushing a few keys.

REFERENCE

B.M. Feinberg and C.A. Franklin (eds.) Social Graphics Bibliography, Washington, Bureau of Social Science Research, 1975.

THE NCES REPORT, 'THE CONDITION OF EDUCATION', 1976'

Mary A. Golladay, National Center for Education Statistics

The Condition of Education is an annual statistical report prepared by the National Center for Education Statistics, Department of Health Education and Welfare, for submission to Congress on March 1 of each year. The first edition was issued in 1975. The report was mandated by Congress in the Education Amendments of 1974. The Amendments created the National Center for Education Statistics, formerly a branch of the U.S. Office of Education, as a separate agency, attached directly to the DHEW Assistant Secretary for Education and serving both the National Institute of Education and the Office of Education as well as many other public and private users of data. They directed NCES to complete several specific studies, examining such problems as impact aid, safe schools, and athletic injuries.

The Amendments require an annual statistical report on the condition of education. NCES was assigned the task of preparing the report, though the content and organization of the report were not precisely specified. The authorizing legislation, directs that the Assistant Secretary for Education, of the Department of Health, Education, and Welfare, submit to Congress on March 1 a report which:

"(A) contains a description of the activities of the Center during the then current fiscal year and a projection of its activities during the succeeding fiscal year;

"(B) sets forth estimates of the cost of the projected activities for such succeeding fiscal year; and

"(C) includes a statistical report on the condition of education in the United States during the two preceding fiscal years and projection, for the three succeeding fiscal years, of estimated statistics related to education in the United States.

In selecting a format and focus for the report, previsous work by the federal statistical system to develop and periodically report indicators of social well being was a major influence. The OMB chartbook, Social Indicators 1973, provided a useful model for style, organization and presentation as well as content and emphasis. A chartbook form of presentation was selected as an efficient means of communicating detailed statistical information in an easily-assimilated format.

Two editions of the Condition of Education have now been issued. This presentation reviews the report as it currently is, raises some of the questions in designating content and choosing presentation, and then considers possible areas of change.

Organization of the Report

The full report contains a statistical report on the condition of education in the United States and a statement of the program and plans of the National Center for Education Statistics. It is to the statistical portion of the report that the following remarks are directed. The report consists of chapters with text and charts, with all tables supporting the charts appearing together in a separate section.

The chartbook portion of the report is presented in two parts. The first part, consisting of three chapters, presents a core of statistics chosen to provide an inclusive, if sketchy, view of education. The second section of five chapters presents more detailed information on selected information on selected topics, with the choice of topics to vary from year to year. This organization was chosen to provide a variety of content in the publication while responding to the demands of an annual publication schedule. Efforts are made to provide comprehensive statistics on educational offerings, participants and outcomes, balanced in coverage overtime but not necessarily in a single year. Special emphasis chapters in 1976 examined the following topics:

Postsecondary Education Participation
Relationships Between Education and Work
Education Personnel
Elementary and Secondary School Finance
International Comparisons of Education in the United States and Other Countries

Each chapter contains descriptive text interspersed with groups of charts. Depending upon the chapter, the text may parallel, complement, or extend the information presented in the charts. A "running commentary" is provided in the form of one-sentence statements accompanying each chart.

Content

The report has the responsibility of informing the Congress as to the condition of education. Selection of content to fulfill this responsibility requires resolution of the antecedent issue of audience definition. The report obviously is intended for persons with some interest in education, but within that large and heterogeneous group it is possible to designate as a primary audience those persons who make decisions which have a direct impact on the content and consumers of educational services. This audience includes representatives of many groups and institutions: Congress and Congressional staff, Federal agencies concerned with education, State education agencies educational institutions, and boards of education. Note that this enumerations leaves out several groups. The reports furthermore is not intended to serve directly the more specialized needs of educational researchers. In summary, the report is a statistical report intended for use not only by statisticians but also by many lay persons and educators who may or may not possess some statistical sophistication.

Given a heterogeneous audience, some decisions were made about content. The report contains many descriptive statistics which can not be considered "indicators": enrollments, degrees, expenditures. It also contains several statistics that are indicators: the high school graduation rate, participation rates in higher education, performance on particular exercises. It was believed that those statistics which describe directly educational services and educational results should necessarily be included, even though they did not meet a formal criterion as indicators. Making this choice of entries required selecting from many available statistical series (and having to approximate desired but non-available series). A core of statistics were selected which include information on participants in education, characteristics of those participants, returns participants receive from schooling, educational offerings and the institutions which offer them, and the costs of providing educational experiences.

The report presents information about education together with statistics describing the larger societal context in which it occurs. It is in the breadth of coverage that this report differs from some other describing education. The report includes statistics on the population, the economy, the labor force, and public opinion on a variety of topics.

The measures used to describe inputs, processes or outputs of education should be sensitive to changes over time and should provide some basis for establishing whether conditions are improving wor worsening. Systems of monetary accounts which trace economic performance provide enviable but as yet unattainable models for reviewing changes impacting on education. In lieu of derived measures, simple time series statistics are used currently. Unfortunately, series data exist for some, but certainly not all, of the variables which might be identified as important. Series data are often readily available to portray characteristics of institutions, yet do not exist to describe persons and their interactions with those institutions.

The report does not sponsor data collection efforts. It is a compilation of published and unpublished descriptive statistics drawn from NCES surveys and many other government and non-government sources. In the 1976 edition, one-third of the entries were from sources outside of NCES. This of course raises the problems which derive from the presentation of non-comparable or not - quite - comparable statistics. The report has been deliberately limited to a presentation of data which have a national base. Some statistics are presented for States or regions, but are provided for the purpose of delineating the condition of education at a national level or the scope of the federal role in responding to educational needs. Disaggregations are thus intended to identify particular areas of change, in order to evaluate impacts of national policies or review those under consideration.

Presentation

The Condition of Education differs from the first OMB Social Indicators publication in several ways. It is instructive to review three aspects of the report in comparison with the earlier OMB report:

(1) graphical presentation

(2) discussion of statistical properties of presented data

(3) analysis and interpretation of data

Graphical presentation. The range of graphic styles chosen for the first two editions has been deliberately limited to ensure (hopefully) accessibility to a large audience which includes laymen and policy makers. Consequently, several candidate presentations have been omitted. The book is comprised of simple bar charts, line graphs and, in the 1976 edition, percentage distributions over time, cumulative distributions over time, and histograms. Missing are growth rates, semi-log scales and regression equations. One might ask whether such self-imposed restrictions are appropriate. Such bland treatment of the data may render the report limited in utility. The statistical sophistication of the general readership is an important unknown here, and perhaps the report has been unnecessarily cautious. While Condition of Education, 1976 did go beyond Social Indicators, 1973, attempts to offer more detailed graphic presentations would seem desirable.

Discussions of Data Characteristics. The Condition of Education has only in special cases presented information describing the characteristics of a data base or the means by which the data were gathered. The 1976 edition contained descriptive discussions of five of the data bases from non-NCES sources used in the report. A more comprehensive set of notes describing the data sources is needed, and more complete discussions of data will be provided in future editions.

Analysis and Interpretation of Data. The Condition of Education is intended to present statistics describing American education in a context which will permit interpretation and contribute to the data base for decision making. Considerable emphasis has been given to presenting current statistics. Such an emphasis runs counter to attempts to conduct sophisticated analyses of data. Trend analysis is possible, but detailed studies will necessarily not be analyses of the latest data.

Aside from the problem of time constraints deriving from data availability and production schedules, plans for the future do call for more analytical report, to the extent that this is possible. It is hoped that some analyses will be conducted specifically for inclusion in the report.

Summary

The Condition of Education is a report still in its infancy. Improvements in data collection and timely reporting, graphical techniques, and analytical capability should all be expected to contribute to improved reports in the future. The process of preparing the report is in itself instructive, identifying those areas where either data gaps or methodological deficiencies impede our ability to report effectively and usefully on the condition of education.

THE OMB REPORT, SOCIAL INDICATORS, 1976

Denis F. Johnston, Office of Management and Budget

"Designers of statistics are indeed philosophers, however unwilling to claim the name, and are fully aware that different aspects of reality can be lit up if alternative sets of concepts are used." --Bertrand de Jouvenel

Social Indicators, 1976 (SI 76) is the second report of its kind to be prepared by the Statistical Policy Division, Office of Management and Budget. It is currently in the final stages of preparation and is scheduled for publication by the Government Printing Office by December, 1976. The general format of this report is similar to that of its predecessor, Social Indicators, 1973 (SI 73) which was issued in February 1974. It again features the graphic presentation, in color, of summary descriptive data on the socio-economic characteristics of the population of the United States, with limited geographic detail but with considerable disaggregation by age, sex, color and other variables.

The contents of SI 76 have been organized into twelve chapters -- an introduction and eleven "social indicator" chapters, as follows:

Introduction (a new chapter)
Chapter 1. Population
Chapter 2. The Family (new)
Chapter 3. Housing
Chapter 4. Social Security and Welfare (new)
Chapter 5. Health and Nutrition
Chapter 6. Public Safety
Chapter 7. Education and Training
Chapter 8. Work
Chapter 9. Income, Wealth, and Expenditures
Chapter 10. Cultural activities, Leisure, and Time Use
Chapter 11. Social Mobility and Participation (new)

The contents of each of the eleven chapters are being presented in three parts: Text and Charts, Statistical Tables, and Technical Notes. The text and charts are further subdivided into sections in accordance with the main topics covered. For example, the population chapter comprises four sections: population growth, population distribution, public perceptions, and international comparisons.1/

Several additional features of the forthcoming report may be mentioned. First, the introduction will provide selected socio-economic data relating to a number of ethnic groups, drawn of necessity from the past three decennial censuses. A few color maps of the U.S., showing similar data by county, will also be included here. Finally, the introduction will provide some guidelines on the use of the report, its organization, and a brief discussion of the quality of the data, organized according to the major sources of data rather than with reference to particular data series. Second, we plan to introduce a few "public perceptions" items in the penultimate Section of each chapter. Nearly all of these items have been selected from the General Social Surveys of the National Opinion Research Center, University of Chicago.2/ Third, we have included selected international comparisons, where available, in the final section of each chapter. Finally, we shall provide an index at the end of the report which will identify materials of related interest that appear in different chapters.

With these general comments aside, it might be useful to devote the remainder of this paper to a discussion of our attempts to respond to (1) some of the criticisms expressed by reviewers of SI 73 and (2) the many suggestions we have received for improving SI 76. These remarks are organized around twelve "problem-areas."3/

(1) Conceptual Organization -- Those who follow social indicator developments are aware of a number of attempts to develop general organizational frameworks for handling the diverse subject-matter to be included in any report of this kind. The organizational structure we have adopted is oriented toward broad areas of "social concern" as these have evolved through the ongoing efforts of the OECD Working Party on Social Indicators.4/ This development effort seeks to specify indicators for nine major concern areas (or "goal areas"): Health, Individual Development through Learning, Employment and Quality of Working Life, Time and Leisure, Personal Economic Situation, Physical Environment, Social Environment, Personal Safety and the Administration of Justice, and Social Opportunity and Participation. Our present organization covers eight of these nine areas in varying degree; only the area of the physical environment is omitted completely.5/

It is probably axiomatic that any principle of organization designed to present some comprehensive set of social indicators leads to boundary problems in handling cross-cutting phenomena and to some danger of "reification" of categories. The chapter organization used in SI 76 accords reasonably well with long-established functional components of our society and enjoys the pragmatic advantage that is also corresponds substantially with a number of established agency jurisdictions within the Federal statistical system. But the objections remain: to treat education separately from work or from cultural activity and even leisure requires the arbitrary placement of data relating to such educationally-significant activities as on-the-job training, reading, attendance at plays, and the like. A more important objection is the fact that this familiar compartmentalization makes it difficult to portray a sense of the flow of experience of individuals at different stages in the life cycle, stemming from their simulataneous involvement in all or most of these "areas of concern." Thus our attempt to depict the "well-being" of individuals remains fragmented.

(2) Analysis and Interpretation -- The old adage that 'a good chart saves a thousand words' is certainly one of the major inspirations for the preparation of a report of this kind, but it fails to answer the two critical questions: what constitutes a good chart and how many words are still required after a thousand have been saved? Several reviewers of SI 73 commented critically on the inadequacy of the text. It was pointed out that the data presented were devoid of any interpretation and that necessary caveats and qualifications were either absent or buried in the Technical Notes. Our response may not fully satisfy these critics. A chart book is a chart book; the inclusion of analysis and interpretation would entail the sacrifice of considerable graphic material and would, more importantly, require changing the basic focus of the report.

We are hopeful, however, that this particular weakness will be overcome by virtue of a separate publication, in the Summer or Fall of 1977, of a special issue of The Annals which will be devoted to a number of expository essays relating to the several chapters of SI 76. Dr. Conrad Taeuber, who served as Chairman of the OMB Advisory Committee on Social Indicators, has agreed to serve as Special Editor of this issue and is currently assembling a group of scholars to carry out this task. Completion of this special issue is not yet assured, but current plans call for the incorporation of much of the data presented in SI 76, so that the readers of this issue will receive a reasonably complete account of the report as a whole, plus a number of interpretive essays.

(3) Scope and Coverage -- Given some upper limit on the size (and cost) of any such report, an obvious trade-off problem arises with respect to adequate coverage of a given topic and the need to treat a broad range of subjects. SI 76, like its predecessor report, opts for a wide range of coverage at the sacrifice of depth. But accepting this decision, the problem remains of making an optimal selection of information; we can only claim that we have tried to do so.

(4) Quality of the Data -- Several of the reviewers of SI 73 expressed strong objection to its failure to convey adequate warnings in regard to the highly variable quality of the data presented or even to provide any awareness of the nature and impact of sampling error. Here again, our response may not satisfy our critics. Our knowledge of the errors associated with data sets bears a strong positive correlation with the quality of these data. We thus encounter the dilemma that any discussion of data quality, especially in a report primarily intended for an audience of non-statisticians, is likely to convey the unfortunate impression that our best data are weakest and our weakest data are best. Of course, nobody proposed that future social indicator reports should attempt to provide a detailed treatment of this complex problem-area; such an effort would be far beyond the capacity of our meager resources in any case. But it was argued that the subject should at least be mentioned, and this we intend to do by providing a brief discussion of the types and sources of error associated with each of the principal types of data sources utilized in the report. This discussion, to be included in the introductory chapter, will necessarily be quite general and elementary, but it will introduce the readers to this problem-area.

(5) Data Presentation -- Most of the criticisms falling under this rubric relate to mis-judgments in graphic presentation, failure to employ appropriate standardization in offering comparisons of data, and the limited use of projections and of cohort time series. We have tried to eliminate the mis-judgments that were generally recognized. For example, we have made greater use of semi-log scales in presenting data of different magnitudes on the same chart, and bowing to convention, we have brought our readers from the second to the first Cartesian quadrant by placing the graph labels for the Y-axis on the left side rather than the right. The more difficult issue posed by the non-comparability of unstandardized summary measures remains to be resolved. The resources available for this project have not permitted any extensive re-working of the data submitted to us by the several agencies. We have of course utilized standardized data, such as mortality rates, when they were made available, but the bulk of the summary measures we present (averages, rates, etc.) remain uncontrolled for the possible effects of changes in the distribution of their components.

We have included more projections than in SI 73 -- including projections of population, households and families, housing demand and labor force. But here also, these projections are those which have previously been developed by other agencies; none has been prepared specifically for this report. Similarly, our hopes to present a number of time series in the form of cohort progressions have been realized to a modest degree with decennial census data on marital status and Current Population Survey data on rates of economic activity and income.

Other serious problems remain. Descriptive indicators may invite invidious comparisons between population groups, particularly in regard to data on crime rates, family instability, and the like. Because the reasons for observed differences are seldom fully understood and, even if known, cannot possibly be explained with the necessary brevity, the presentation of such information in chart form is subject to the objection that it may lend support to distorted perceptions or unwarranted conclusions based on prejudice rather than fact. Of course, all social information is subject to the danger of misinterpretation. That danger must be weighed against the dangers arising from the withholding of pertinent information because it is controversial or provocative.

(6) Descriptive versus Explanatory Indicators -- The data presented in SI 76, like those in SI 73, are almost exclusively descriptive in character. The only complex or analytically derived measures which are included are those well established measures, such as estimates of average life expectancy, which are conventionally employed in

similar reports. The cross-tabulations which are shown are mostly two-dimensionsal. No attempt has been made to relate different sets of data in a statistical sense or to report the findings of the more elaborate multivariate analyses of social phenomena which have been carried out. In short, to employ Richard Stone's terminology, SI 76 is designed to satisfy (or rather to arouse) curiosity but not to provide "understanding" or "prescriptions for action."[6/]

(7) Normative Considerations -- To some of the pioneers in the field, the very term "social indicators" implies a normative element or focus. In fact, it has been argued that it is the normative element which distinguishes indicators from other social statistics.[7/] But current practitioners in the field, particularly those who regard social indicator development as an integral part of normal social science research, recognize a basic confusion in this argument: indicators, they point out, are value-neutral; the normative significance of any datum lies, as does beauty, in the eyes of the beholder. In other words, the determination that a given statistical trend or measure indicates that something has gotten "better" or "worse" is entirely a function of the valuational perspective of the observer. Granting that the producer of social indicators is engaged in a purely technical, value-neutral enterprise, governed by the norms (values?) of objectivity, the fact remains that the producer of a social indicator report cannot claim such neutrality in quite the same sense. His primary task is the judicious selection and presentation of information relating to a number of social concerns.

In practical terms, this means that judgment, reflecting some set of values, must be exercised in carrying out both the selection of data and in devising some mode of presentation. To choose a controversial example, SI 76 will show, in the family chapter, the rise in the divorce rate in the United States. We may presume that some observers will recognize this trend as "bad" while others may counter with the argument that it also reflects some "goods." Such assertions do not necessarily imply confusion between neutral "facts" and valuational interpretations; they are elliptical statements and obviously refer to the underlying phenomenon and its associated social and psychological conditions. Divorce is "bad" because it commonly entails family disruption, psychic strains, legal costs, and the like. It may also be "good" insofar as it reflects individual self-reliance, flexibility and a willingness to seek improved human relationships. Neither interpretation fully contradicts the other, but each one clearly moves us well beyond the basic measure we started with.

Which bring us, finally, to the point: the decision to show the trend in the divorce rate is not value-neutral. It is judgmental and reflects what is presumed to be a widely shared concern with one of the basic values of our society -- family stability. The likelihood that this trend will prompt divergent interpretations is an appraisal that can be made of most, if not all of the data presented in the report. As a corollary point, one can argue that a basic objective of reports of this kind is to prompt reasoned discussions or consideration of the issues and problems which are apparent in the several areas of concern while seeking to raise the level of factual information on which such consideration may be based.

(8) Disaggregation -- One of the more penetrating comments on SI 73 was Natalie Ramsøy's observations concerning the almost universal and uncritical employment of age-sex-race classifications in disaggregating most of the data presented.[8/] In brief, she argues that the explanatory significance of these variables varies widely among different social phenomena and may in fact be largely irrelevant with respect to some of them. By showing such disaggregations uniformly, we are, in effect, encouraging the notion that these variables are uniformly relevant while at the same time failing to disclose (or masking) other relationships which may be more significant. We have tried to introduce other breakdowns when they were available to us and we share the view that disaggregations by socio-economic status, life cycle stages, occupation, and education should be included more extensively than they have been up to now. However, age, sex, and racial differences should still be included with respect to most of the socio-economic data shown in a report of this kind. As Otis Dudley Duncan has pointed out, normative considerations dictate such disaggregations because of the need to describe the relative status of these groups, quite apart from the explanatory significance of these background characteristics. While this argument does not invalidate Ramsøy's objections, it does indicate why such disaggregations will continue to be widely employed.

A different problem has to do with space limitations. SI 76 has sometimes displayed the top row of a table of data (measures relating to the total population through time) on one chart and the right-hand column (measures relating to subgroups for the latest period of observation) on a second chart. This was done in SI 73 as well and is perhaps the best compromise between showing all the detail available in a table and showing only total values.

(9) Perceptual or Subjective Indicators -- Our original intent was to organize the data in each chapter according to a scheme proposed by Wolfgang Zapf, whereby three basic types of social indicators are recognized: indicators of system performance (primarily institutional resource inputs and programmatic output measures); indicators of well-being (objective measures of effects or outcomes); and indicators of satisfaction (data reflecting popular perceptions or feelings about aspects of their condition or situation, prospects, etc.)[9/] This plan failed; we were unable to force the diverse kinds of data available to us into this format. But the underlying concept remains valid as a delineation of the types of indicators which are called for in providing a basis for a comprehensive assessment of "well-being" in the broadest sense.

Given that objective, the provision of objective measures of different aspects of well-being is clearly necessary and clearly insufficient. The need to supplement such data with subjective meaures is supported by W. I. Thomas's dictum that it is not the objective situation that serves to explain human action, but how that situation is perceived by the actor. Of course, Thomas was concerned with research objectives; he sought to deepen our understanding of the determinants of human behavior. But if subjective measures are called for in such a case, they are just as essential in program evaluation, where public perceptions and reactions are bound to affect program outcomes, and in public assessments of the condition of the society, where changing values and aspirations often presage important social changes. Hence our concern to include such data in every chapter, despite their commonly acknowledged limitations and difficulties of interpretation, reflects our conviction that data on public perceptions constitute an essential component of social indicators.[10/]

(10) Distributive versus Collective Well-being -- As stated in the introduction to SI 73, the choice of indicators was based upon two main criteria, the first of which was that the indicators should measure the well-being of individuals (or families) rather than that of institutional entities. Given this intended focus, SI 73 was faulted for omitting data on family status and on social welfare -- two omissions which have been remedied to some extent in SI 76. But a more profound issue was raised in this regard: can we accept the implicit assumption that societal "well-being" is merely the sum of the well-being of its individual members? This form of psychological reductionism does not sit well with many social scientists, especially when we consider the need, in all societies, to restrict individual behavior in many ways in order to optimize the well-being of society as a whole. There is therefore an arguable need to include information on the performance characteristics (not to say "well-being") of major functional sectors of society, expressed in terms of both inputs and outputs. SI 76 has made a modest start in this direction, but at the risk of losing a distinctive focus on individual well-being.

(11) Input versus Output -- The second criterion for the selection of social indicators, as expressed in the introduction to SI 73, was that the measures selected should reflect "outputs" or "results" rather than "inputs." If the preceding argument has merit, the omission of input measures impairs the effectiveness of the report as a tool for public assessment of the condition of the society. Furthermore, the application of the distinction between "input" and "output" in non-economic areas gives rise to serious problems. What is educational achievement (i.e., learning) an output of? What about an individual's health status? Even the familiar statistic relating to "years of school completed" is clearly an output of the educational system and an input to the labor force and to other sectors of the society.

Such questions are quickly resolved once agreed-upon sub-system boundaries have been established; the distinction is a function of one's perspective and research objectives. But that is the point at issue: the imposition of such a perspective, which is essential to developing a systems-analytic conceptualization of society, exceeds the limited aims of SI 76. By adopting a looser framework, we hope to offer the readers of the report greater degrees of freedom in approaching its contents from a variety of perspectives.

(12) The Quality of LIfe -- If one of the ultimate objectives of the "social indicators movement" is to enhance our ability to assess the quality of our lives, one of the sobering insights acquired in preparing a national social indicator report is an awareness of the multi-dimensionality of this elusive quality. The effort to develop a single composite index of the quality of life retains its supporters, despite Bertram Gross's earlier warning of the "philistinism" implicit in the imposition of a single metric to such vital components as health status, marital happiness, feelings of efficacy, job satisfaction, adequate housing and, of course, income security. But if the objective of devising a single index of life quality seems far fetched, the aim of providing, in a single report, a selection of descriptive information to assist readers in developing a more adequate assessment on the basis of their own values and priorities is not invalidated.

Conclusion

It is OMB's judgment that, on balance, the forthcoming social indicators report is a distinct improvement over its predecessor. If that claim turns out to be warranted, those who prepared the first report will merit a large share of the credit, since they have greatly facilitated the preparation of the second. It must also be admitted that some of the changes and additions may, upon review, be assessed as retrogressive rather than progressive. It is clear that no "great leap forward" has been achieved. Such advances as have been made have instead been by way of what O. D. Duncan might term "pragmatic incrementalism."

Another difficulty remains in preparing reports of this kind -- continuing uncertainty as to its actual users and usage, as distinct from those which were envisioned originally. We know that the report, like its predecessor, will provide a large number of government officials and their staffs with summary information on a variety of topics relating to the country as a whole and presented in a uniform manner. It will, we hope, serve a similar purpose for those "micro decision-makers" among the general public who wish to acquire a similar "macro" perspective. We also know that graphic presentations of this kind (as in the new journal, STATUS) are found useful in providing a "quick fix" on a given topic under circumstances which preclude detailed investigation. Those technical experts who shudder at the thought that major decisions may be arrived at on the basis of a quick glance at a few graphs should perhaps be reassured that a great many other elements enter into the "fluid drive" between informational inputs and the act of decision. Furthermore, most decision-makers, both public and private, are busy generalists, not busy specialists. For them, a quick fix in the form of summary statistics may often be an important corrective to hunches and intuition. We have also learned, in a spotty manner, that librarians find these reports to be extremely useful in guiding students toward a few general insights on the subjects covered. This may not be cheerful news to those who envision reports of this kind as powerful vehicles for influencing major policy decisions, but the teachers among us may be gratified.

Finally, we can assert that the report is not intended to serve the research needs of statisticians, social scientists and similar experts, except insofar as they, too, may perceive a need for a handy summary of descriptive data, particularly in areas with which they are not deeply familiar. We can also assert, with confidence, that most of the charts in the report raise more questions than they answer, prompt a variety of conflicting interpretations and promote a strong demand for additional information -- characteristics they share with other forms of research reporting.

Footnotes

1/The bulk of the data for this report has been assembled with the assistance of the members of the Interagency Committee on Social Indicators, chaired by the author, and comprising the following persons: Jack Blacksin (Internal Revenue Service), Arnold H. Diamond (Dept. of Housing and Urban Development), Jacob Feldman (National Center for Health Statistics, DHEW), Walton Francis Dept. of Health, Education, and Welfare), Iris Garfield (National Center for Education Statistics, DHEW), Harold Goldblatt (Dept. of Housing and Urban Development), Max Jordan (Dept. of Agriculture), Sue Lindgren (Law Enforcement Assistance Administration), Alfred Skolnik (Social Security Administration, DHEW), Robert Stein (Bureau of Labor Statistics), Gooloo Wunderlich (Public Health Service, DHEW), Meyer Zitter (Bureau of the Census), and Paul Zolbe (Federal Bureau of Investigation). The report is being prepared in the Statistical Policy Division, OMB, by the author with the assistance of Marian Altman and Tobia Bressler who are on detail from the Bureau of the Census. It is being prepared in final form for printing by GPO by the staff of the Publications Services Division, Bureau of the Census.

2/The General Social Surveys conducted by NORC began in 1972 and are scheduled for annual repetition (of most items) through 1978. Thus the time series now available (1972 through 1975) are brief, but the data will eventually yield a picture of trends through the 1970's.

3/This discussion draws heavily upon the reactions of a panel of social scientists and statisticians which was assembled to review SI 73 in late February, 1974. Their comments are presented in Roxann A. Van Dusen (ed.), Social Indicators, 1973: A Review Symposium (Washington, D.C.: Social Science Research Council, Center for Coordination of Research on Social Indicators, 1974.) It aso reflects some of the comments made by individual members of the OMB Advisory Committee on Social Indicators, which included Conrad Taeuber (Chairman), John H. Aiken, Daniel Bell, Albert D. Biderman, Angus Campbell, David Christian, Otis Dudley Duncan, Jack Elinson, David A. Goslin, Abbott L. Ferriss, Lloyd A. Free, Harvey A. Garn, Robert B. Hill, F. Thomas Juster, Ida C. Merriam, Graham T. T. Molitor, Milton Moss, Daniel Patrick Moynihan, Roberto Olivas, Robert Parke, Nestor E. Terleckyj, Daniel B. Tunstall, Ralph R. Widner, Willard W. Wirtz, and Marvin E. Wolfgang.

4/Organization for Economic Cooperation and Development (OECD), List of Social Concerns Common to Most OECD Countries (Paris: OECD, 1973). The nine areas listed above reflect current modifications of this original list of concerns. The modified listing and related discussion are contained in a progress report on Phase II of the activies of the OECD Working Party on Social Indicators (publication forthcoming.) Most of the member countries of OECD have issued one or more "social indicator" reports, including Canada, France, Japan, the Netherlands, Norway, Spain, Sweden, the United Kingdom, and West Germany.

5/The eleven chapters of SI 76 also correspond quite closely with the eleven major "sub-systems" currently identified in the United Nations Statistical Office's "System of Social and Demographic Statistics" (SSDS). This system, which was developed under the general direction of Professor Richard Stone of Cambridge University, is most fully described in UNSO, Towards a System of Social and Demographic Statistics (New York: United Nations, ST/ESA/STAT/SER.F/18, 1975).

6/Professor Stone's fuller statement is as follows: "...Social indicators relate to some area of social concern and they may serve the purposes of curiosity, understanding or action. They may take the form of simple data series or they may be synthetic series obtained by applying a greater or lesser lesser amount of processing to data series... Social indicators form a subset of the data series and constructs actually or potentially available and are thus distinguished from other statistics only by their suitability and relevance for one of the purposes mentioned." Ibid., p.28.

7/Both the normative and the "program evaluation" concerns are evident in the definitions propounded by the principal authors of two of the more influential efforts in the field. First, RAymond Bauer's statement: "This volume as a whole is devoted to the topic of social indicators -- statistics, statistical series, and all other forms of evidence -- that enable us to assess where we stand and are going with respect to our values and goals, and to evaluate specific programs and determine their impact." Raymond A. Bauer (ed.), Social Indicators (Cambridge, Mass.: The M.I.T. Press, 1966) p.1. Second, Mancur Olson's statement: "A social indicator, as the term is used here, may be defined to be a statistic of direct normative interest which facilitates concise, comprehensive and balanced judgments about the condition of major aspects of a society. It is in all cases a direct measure of welfare and is subject to the interpretation that, if it changes in the 'right' direction, while other things remain equal, things have gotten better, or people are 'better off.' Thus statistics on the number of doctors or policemen could not be social indicators, whereas figures on health or crime rates could be." U.S. Dept. of Health, Education, and Welfare, Toward a Social Report (Washington, D.C.: Government Printing Office, 1969) p. 97.

8/ In Van Dusen (ed.), Op. Cit., pp. 46f.

9/ In Van Dusen (ed.), Op. Cit., 26.

10/ On this issue, the following comment by John Tukey is apposite: "...It is often much worse to have a good measurement of the wrong thing -- especially when, as is so often the case, the wrong thing will IN FACT be used as an indicator of the right thing -- than to have poor measurements of the right thing." John W. Tukey, "Methodology, and the Statisticaian's Responsibility for Both Accuracy and Relevance." Presented at the General Methodology Lecture at the 135th Annual Meeting of the American Statistical Association in Atlanta, Georgia, August 1975. Reprinted in the Statistical Reporter, No. 76-13 (July 1976) pp. 253-262.

DISCUSSION

Angus Campbell, Institute for Social Research, The University of Michigan

Like many of you I have a strong interest in descriptive statistics and I found the reports the speakers are responsible for full of fascinating data. Perhaps a somewhat different selection of tables might have been chosen by people with different interests but as a collection of general interest data, I can't fault it. Some people might have preferred more textual material, more analytic interpretation of the data. The British publication Social Trends, for example, uses more commentary. But this takes up space and reduces the total display of data. It may also discourage readership--some people find it more congenial to scan through a succession of tables than to read someone's explanation of what they mean. I think I prefer the format these reports have followed, many tables with a minimum of comment.

I propose to restrict my comments to some observations as to the potential social value these reports have. You are certainly aware that social scientists are under some pressure these days to show the social relevance of the research they are doing, to prove that the utility of their research justifies the cost. I think myself that there is a certain amount of know-nothing, anti-intellectualism mixed up in this but some of it is coming from rather high places and we would be prudent to take the question seriously.

Denis Johnston says that Social Indicators 76 is designed to satisfy (or arouse) curiosity but not to provide "prescriptions for action" or "understanding." He has borrowed this terminology from Richard Stone at Cambridge University. Let me first consider curiosity.

I have no doubt that the data in these reports will stimulate the curiosity of many readers. I find, for example, in Health, United States, 1975, that black people in this country are less likely to lose their teeth than white people. This strikes me as very curious in view of the fact that loss of teeth is quite clearly associated with income level and that, on the average, black people visit a dentist only about half as often as white people do. That bit of information will remain a curiosity to me; it is not likely to affect any research I do or any decision I may be party to. But to an epidemiologist interested in dental caries it would certainly be much more than a curiosity, it would be close to the heart of his professional interests.

On the other hand, I see from The Condition of Education, 1976, that the percent of population participating in adult education activities has almost doubled between 1967 and 1975. And that is considerably more than a curiosity to me because it has far-reaching implications for my profession and the institution which employs me. The moral is, I suggest, that what is one person's curiosity is another person's obsession.

Now what about using these data as "prescriptions for action"? I think Denis Johnston is wise in pointing out that there are many considerations that intervene between the informational inputs that flow out of these reports and any subsequent acts of decision. There is, in my view, a good deal of naivete abroad in the land about how research gets turned into action. The picture of an anxious public administrator waiting for the one right table that will solve all his problems is surely not very true to life. And the expectations of some high level public officials that social science should produce an effective "fix" for such social pathologies as crime and the fragmentation of families is also not very realistic. It must also be admitted that some of our social science colleagues have got a little overheated in their belief that they had the one right answer in such programs as the War on Poverty. And some of the predications they launched on the public five or ten years ago do not look very good in retrospect.

There is no doubt that those of us in the social sciences have a great deal to feel modest about. At the same time there is clear evidence that policy-makers are listening to what the social scientists have to say. The Institute for Social Research has recently completed a study of the utilization of social science data by high-ranking officials in the executive offices of the federal government. This study, which was under the direction of Nathan Caplan, interviewed 204 people just below cabinet rank. These people were asked if they had at any point made use of social science information, other than specifically economic information, in their decisions as government officials. Nine out of ten said they had and they reported a total of 575 specific instances of such use. There was some duplication in these reports, leaving some 450 separate cases of the use of data growing out of social science research, excluding economic data.

Is this figure impressively high or disappointingly low? Perhaps some of both--it at least shows that these people were aware of some the output of social scientists. It would be interesting to know if this awareness is growing but that we cannot know from a single study. None of the reported instances of data use was crashingly dramatic. Many of them resulted from specific studies the agencies themselves had sponsored; half of them dealt with inhouse problems. However, a large fraction of them were drawn from survey research findings, demographic research and social statistics of one kind or another. Nearly all of these people expressed an interest in data which we classified as falling within the realm of social indicators.

It is interesting to see the priority these people gave to the uses of social science data, the order of utilities they saw in these data. They ordered their usefulness as follows:

1. Sensitizing policy-makers to social needs,
2. Evaluation of ongoing programs,
3. Structuring alternative policies,
4. Implementation of programs,
5. Justification of policy decisions,
6. Providing a basis for choosing among policy alternatives.

I don't cite the data from the Caplan study to demonstrate how important social scientists, other than economists, have become in public affairs--I don't believe they demonstrate that. They may be useful, however as an antidote to the sense of despair some social scientists seem to feel, the feeling that no one is paying any attention to what they are doing. On the contrary, these federal executives were very much interested in what was going on in the world of social science and concerned with learning more about it.

Now let me return to Denis Johnston's statement and ask whether social indicators can be expected to provide "understanding." I would think this depends on what we mean by understanding." In 1960 Rensis Likert gave the presidential address to this Association; his title was "The Dual Function of Statistics." His argument was that data are of two kinds, those that describe the state of an organism or system and those that explain its nature. Johnston is saying that social indicators will describe the state of society but will not provide an understanding of its functioning, and in this he is probably right. I would point out, however, that we are very unlikely to achieve an understanding of the nature of our society's functioning until we know the basic descriptive facts about the state of our society.

You are well aware that there is a certain impatiience in some circles with what is called "measurement without theory." People with this point of view would like to have a model or theory that explains the nature of society before one undertakes to take measurements of the state of society. This argument is reviewed by Dudley Duncan in his Russell Sage monograph, Toward Social Reporting, and I cannot resist a brief quotation from it.

> "To sum up, if not caricature, the two positions: The 'theorist' says, 'Let us think long and hard about what we want to measure and why. Then we will feel confident about what ought to be done by the way of making observations.' The 'inductivist' responds 'Let us see if we can measure something, for whatever reason, and standardize our measments so that we achieve an acceptable level of reliability. Then let us study how the quantity being measured behaves. If we can figure that out, we will have come to understand why we made the measment in the first place.
>
> It is suggested that the history of science will provide instances in which each of these approaches was successful Anything that works cannot be dogmatically rejected. The strategic question at a given moment, of course, is where to place one's bets. My own assessment at this time is that those who have approached the problem of social reporting with the strongest theoretical presuppositions have possibly made the least impressive contribution thus far."

If we take a less exalted definition of understanding than I think Denis intended I think it is clear that social indicators do have an important influence on what decision-makers and the public at large understand social reality to be. Descriptive data can change their perceptions, their cognitive map of society, and the consequences may be far reaching.

Consider for example the change in the racial situation in this country since World War II. I would not argue that the development of descriptive data about the status of black people was solely responsible for this change but I believe firmly that it contributed significantly to it. When I began teaching social psychology in 1936 there was virtually nothing I could assign my students on the realities of the caste system in the United States. To these white, middle-class young people the black population was virtually invisible. Toward the end of this decade there came a series of books describing the lives of black people, Franklin Frazier's The Negro Family in 1939, John Dollard's Caste and Class in a Southern Town, the American Council on Education series on black youth, and in 1944 Gunnar Myrdal's The American Dilemma.

What these books did was to heighten the awareness of the book-reading public to the facts of life about black people, to the racial system of caste and class. In the modern idiom, they were "consciousness raiding." Or to use the language of our study of federal executives, they "sensitized policy-makers to social needs."

I think the same kind of case can be made concerning the change in public policy toward poverty. Fifty years ago poverty was generally understood as a failure of will or a weakness of moral character, people in poverty were thought to be to blame for their condition, and the accepted solution was for these misled people to straighten up and go to work. Some people still see poverty in those terms. But a great deal of information about people in poverty has been publicized in this country in the last 25 years and we have moved to a different perception of the realities of being poor, a realization that very few of the poor are able-bodied men who are avoiding work and that most of the poor are trapped in situations from which they have very little possibility of escape. Because this change in understanding has come about, it is now possible for the Congress to give serious consideration to legislation to guarantee a minimum annual income, a proposal which would have been unthinkable only a few years ago.

In my view the major value of the kinds of statistics we see in these publications is that they restructure people's view of the world; they introduce reality into their perceptions of society. They may also occasionally provide a piece of information that is specifically relevant to some action decision but more importantly they provide what Albert Biderman has called "enlightenment."

I want to make one additional observation about Denis Johnston's paper. I want to compliment him on his courage in supplementing the objective measures of well-being he will include in Social Indicators 76 with subjective measures of satisfactions and other aspects of life experience. I should point out in this connection that The Condition of Education also presents a number of such measures. I say "courage" because, as you well know, there are a good many statisticians and economists who cannot feel comfortable with measures that do not have the qualities of a ratio scale, with a true zero point and equal intervals For them subjective measures are simply "too subjective" and they would far rather use some surrogate whose measurement qualities they have confidence in.

Unhappily, as John Tukey has recently pointed out, "It is often much worse to have a good measurement of the wrong thing than to have poor measurement of the right thing--especially when, as is so often the case, the wrong thing will in fact be used as an indicator of the right thing." Nowhere is Tukey's observation more pointedly relevant than in current efforts to measure quality of life. Despite what is to me the obvious fact that the quality of a person's life is known only to him through his personal experience, there persists an obstinate conviction that we can get a truer measure of this experience by counting that person's income and savings than we can by asking him how his life feels to him.

I think Denis is absolutely right in stating that "one of the ultimate objectives of the social indicator movement is to enhance our ability to assess the quality of our lives and that it is "far-fetched" to imagine that we can accomplish this with any single composite index. What we obviously need is continuing development of both objective and subjective measures and a determined effort to discover how they are related to each other and how they both may be optimally used to help understand the changes which are taking place in American society.

I recently came upon a quotation which I think is particularly apt to our discussion today. It reads, "If we could first know where we are and whither we are tending, we could better judge what to do and how to do it." These lines were spoken by Abraham Lincoln in Springfield, Illinois, on June 16, 1858. He was not talking about social indicators but his observation is as appropriate to our time as it was to his.

SECONDARY ANALYSIS OF PUBLISHED RESEARCH IN THE BEHAVIORAL SCIENCES[1]

Leroy Wolins, Iowa State University

Least-square regression procedures, including analysis of covariance and partial correlation, may be thought of as directed toward two conceptually different goals. The first is prediction. In this case the purpose is to predict future events from earlier ones where there is little or no concern about underlying variables or what causes what. The predictor variables represent nothing other than the operationally defined measurement procedures. A typical example of this is the prediction of Grade-Point-Average from test information.

The second goal is more theoretical. The purpose is to attribute causation. In this case the variables represent constructs and the analysis is directed toward answering questions about these constructs rather than the operationally defined measures of them.

When these predictor variables are orthogonal, these regression procedures tend to be robust and informative. When these predictor variables are not orthogonal the analysis for the second goal is likely to be misleading for a variety of reasons. One reason such analyses are misleading has been well documented in the statistical literature (Cochran, 1970; DeGracie and Fuller, 1972) and in the literature in other fields. Measurement error in the predictor variables tends to produce significant partial regression coefficients for correlated measures not necessarily because these measures are tapping different constructs but possibly because each of these measures tap the same construct but with error. As a result of this error a composite made up of these fallible measures will predict the dependent variable better than a single variable will.

However, measurement error in predictor variables is not the only reason the analysis for this second goal may be misleading. Other reasons for misleading results are less familiar but can be no less important. Many of these can be divided into two general categories: non-addivity or lack of interval measurement and irrelevant variance in the predictors.

In regard to non-additivity, imagine an analysis of covariance situation with two groups, men and women in an academic setting, a dependent variable, Salary, and a covariate, Number of Publications. Suppose it happened that at the lower salary levels that Salary was highly dependent on Publications but at the higher salary levels there was little relationship. That is, the relationship between number of publications (X) and salary (Y) is monotonic but negatively accelerated. Also suppose that males exceed the females both in terms of Salary and Publications but all the data lies on a single curved line. Now if you fit these data using the usual covariance analysis, the partial regression coefficient for Sex, controlling for Publications will be significant, indicating to the naive researcher that males are paid more than females. That is, even if sex was irrelevant in nature, the conventional covariance analysis would produce the semblance of an effect. That semblance is attributable entirely to the nonadditive character of the results, a character which is not reflected by the usual analysis.

Of course, such systematic non-addivity is not likely to occur in the social sciences and one reason it is not likely to occur is that the intervals in our scales are arbitrary in size relative to the construct to which they are directed. For a faculty member in an academic institution, Publications is an indicator of a vaguely defined construct, the extent and the quality of an individual's research. Although we may imagine individual faculty members have a true and invariant location on this conceptual dimension, we have no way of knowing how this indicator is related to this dimension. The dependent variable, Salary, should also be functionally related to this true variable if the indicator of the true variable is a proper control variable. However, the only basis we have for determining whether the dependent variable is related to this true dimension <u>in the same way</u> as that dimensions indicator (Publications) is by evaluating the linearity of the relationship between Publications and Salary.

In addition to non-addivity, most variables used in the social sciences reflect more (or less) than their target construct. For example, it seems perfectly reasonable to suppose that Number of Publications is an indicator of the construct, Research Productivity. However, it is equally evident that all publications are not equal. I believe administrators in universities, and committees assigned the task of making decisions about promotions and salary increases make serious efforts to judge the merit of each publication as well as counting papers or articles. As a result, if a group of faculty were evaluated in terms of number of publications for two adjacent five year periods, one might find a high correlation. However this high correlation might represent, in part, the fact that individuals tend to publish in the same area and that certain areas require large scale, involved research whereas in other areas short research efforts are publishable.

This variability that is reliably measured by these control variables which is not reflected in the dependent variable, does exactly the same thing measurement error does but it is much more difficult to accommodate.

This brings us to the data I have at hand.

In the May 1975 issue of Science there appeared an article entitled "Sex Differentials in the Academic Reward System". The authors, Alan E. Bayer and Helen S. Astin, have been extremely cooperative in every way. They not only sent me their data but have provided other information when I requested it and admonitions concerning various intracies about the data I did not request but should have. This kind of cooperation deserves special commendation, especially since they knew my purpose, criticism, and other authors whose data I requested tended to beg off, dismiss the matter perfunctorily, or not respond at all. I do intend to pursue the problems with these other authors but today unless there is extra time, I will confine my comments to the Bayer-Astin article.

Table 1 contains Salary data based on groups homogeneous with respect to Sex, Academic Rank and Departmental Affiliation. These data are based on about 2000 Males and 2000 Females who filled in their questionnaires properly. Since fewer females than males occur in the academic population, females were sampled more intensively than the males in order to obtain approximately equal numbers of each. This explains some of the peculiar results in Table 2. In the last two columns of Table 2 are the numbers for each sex occuring at each Rank and Departmental Affiliation. For example, females tend to predominate at the lower ranks for this sample whereas males tend to predominant at all ranks in the academic population.

In Figure 1, I have plotted the mean against the standard deviation of salary for each group with data for twenty-five or more individuals. It is quite evident from this figure that these two statistics are not independent which suggests non-addivity. A logarithmic transformation is suggested by the fact that merit increases seem to be based on a percentage of previous salary. Also I tried a square-root transformation and it did not work as well as the logarithmic transformation. In Figure 2 is depicted the same data expressed on the log scale. The apparent relationship between $\bar{X}$ and S is considerably reduced on the log scale.

In Figure 3, only groups containing at least 25 males and females at each Rank and Departmental Affiliation are depicted. For example, for $R = 1$ and $D = 9$, only the females have more than 25 cases so this group is not depicted in Figure 3. From the Figure, it is also evident that all three variables, Rank, Departmental Affiliation and Sex are related to Salary. Moreover it appears that the sex differential increases with salary.

In Figure 4 is depicted the same data on the log scale. Here the differential appears more nearly constant.

(By the way, the two outliers are for the "Health" group. For the males about 80% hold doctorates, for the females about 30% hold doctorates. For the males more than 80% of these doctorates are professional degrees where less than 50% are professional degrees for the females. These results are similar for all three health groups but less marked for the full professors in this health group.)

From these results it appears evident that we are better off, from a statistical point of view if further analysis is based on log salary rather than salary.

In Figure 5, log salary is plotted against number of articles published. Number of articles published is a coded variable but is a monotonic function of the actual number of publications. In this Figure there appears to be six outliers. All six are affiliated with departments of Biology. Except for these six outliers, which appear to reward the females more than the males, no sex differential is evident.

In Figure 6, the number of articles published for Males and Females are plotted. Again it is evident there is a sizable differential favoring the Males for most ranks and departmental affiliations.

In Figure 7 is plotted Log Dollars against Number of Books. Like Number of Publications, this is a coded variable. The three outliers are for the Male "Health" groups. These groups as all ready noted, contain disproprortionate numbers of Ph.D.'s and professional doctorates as compared to females in these "Health" groups comparable in rank.

Bayer and Astin, from their regression analysis, attribute a salary differential from $600 to $1000 per year to sex, favoring males. It is my contention that this result would be expected for this type of analysis because the independent variables used for control are in some sense falliable or the dependent variable is not linearly related to them, or both. The data I have presented seem to indicate people of the same Academic Rank, who publish the same amount, receive the same pay.

As made clear by Bayer and Astin, and very evident from the data presented here, there are sex differences. However, it is not clear that the institution responsible for these differences is the academic one. For example, I note that males at all ranks have more children than females but this differential increases with increasing rank. This suggests that children impede the progress of females more so than males. On the other hand, it appears that males and females report that their academic careers were interrupted about equally often at the lower ranks but males at the full professor rank report more career interruptions than their female counterparts, probably because of military service during World War II.

Although these data do not support a sex differential in the academic reward system, it is essential to understand that these data do not indicate that men and women are equally rewarded either. For example, if one chose to use number of publications as the dependent variable and regressed this variable on the remaining ones, including salary, it is likely that the sex variable would be significant, erroneously indicating men publish more than women when salary etc., is controlled. By exploitating this regression phenomena, one can use data to support whatever position one wishes, under null conditions. Of course, if a strong sex effect were present, then this regression procedure would find it, however, so would this rather casual approach used here.

Bayer and Astin claim these estimates of the sex differential are under estimates for various reasons. I claim they are overestimates for other reasons. In a sense, we agree that these are not good estimates. On these bases I claim analysis of data like this should not exceed in complexity what I have done here. If clear sex differentials are not evident from examining the means for large homogeneous groups, then the data are not adequate for inferring a differential. Seeking a precise probability estimate or an unbiased estimate of a difference for data like this is futile. There is no reasonable way to obtain such precision given fallible measurement.

The reasons that such analyses are never appropriate to adjust for group differences on covariates are as follows:

1) Social science variable do not have interval properties. As a result the construct measured by the covariate may not be functionally related to the true measure in the same way as the measure of the same construct by the dependent variable.
2) The dependent variable may result, in part, from variables not measured by the covariates. For these data it could be that women or men perform better in the classroom and this fact is reflected in the salary variable but not in the covariates.
3) The covariates are fallible measure of their target constructs. As previously indicated number of publications is not likely to reflect completely an individuals research productivity. Research Productivity maybe more completely reflected in salary because administrators or committees may actually scrutinize and evaluate publications.
4) Additivity, particularly linearity, can not be assumed for variables of the kind used in social science investigations. Linear models in the social sciences are used to roughly fit monotonic relationships. However the ad hoc measurement procedures do not allow one to put much faith in a linear model really fitting data. When two groups differ on a covariate and the relationship between the variate and covariate is non-linear, some degree of bias in the results of the covariance adjustment is bound to occur.

What Bayer and Astin have shown from their regression analysis is that one can predict salary with sex as a variable better than if one ignores sex. However it is not clear that this increased prediction is due to sex per se or because these other variables which also reflect sex differences do so fallibly. Since number of publications etc. fallibly reflect research productivity etc., it can not be clear from these data that women are paid less because they are women or because they publish less.

---

[1]Supported by National Institute of Education Contract Number NIE-C-74-0115.

REFERENCES

Bayer, Alan E. and Helen S. Astin. (1975). "Sex differentials in the academic reward system". Science 188: 796-802.

Cochran, W. G. (1970). "Some effects of errors of measurement on multiple correlation". Journal of the American Statistical Association 65: 22-34.

DeGracie, J. S. and W. A. Fuller. (1972). "Estimation of the slope and analysis of covariance when the concomitant variable is measured with error". Journal of the American Statistical Association 67: 930-937.

Figure 1. Plot of $\bar{X}$ and S for the original scale, Dollars

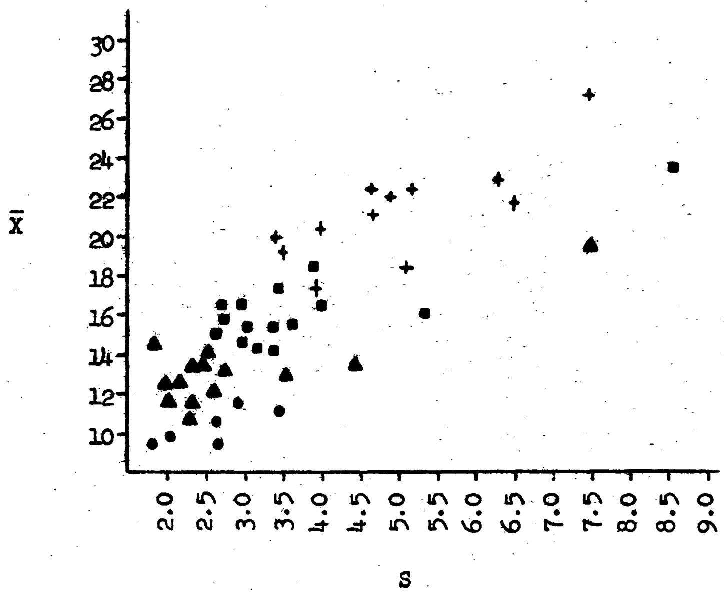


Figure 2. Plot of $\bar{X}$ and S for the transformed scale, Log Dollars

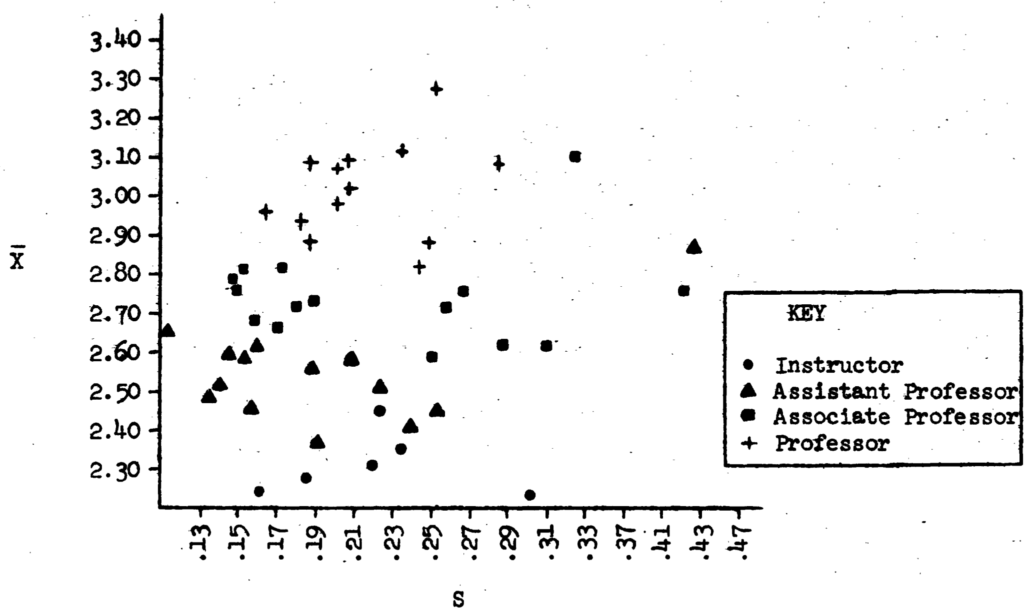


Figure 3. Mean salary (original scale) of Males (M) and Females (F) in the same field and holding the same academic rank

Figure 4. Mean salary (log scale) of Males (M) and Females (F) in the same field and holding the same academic rank

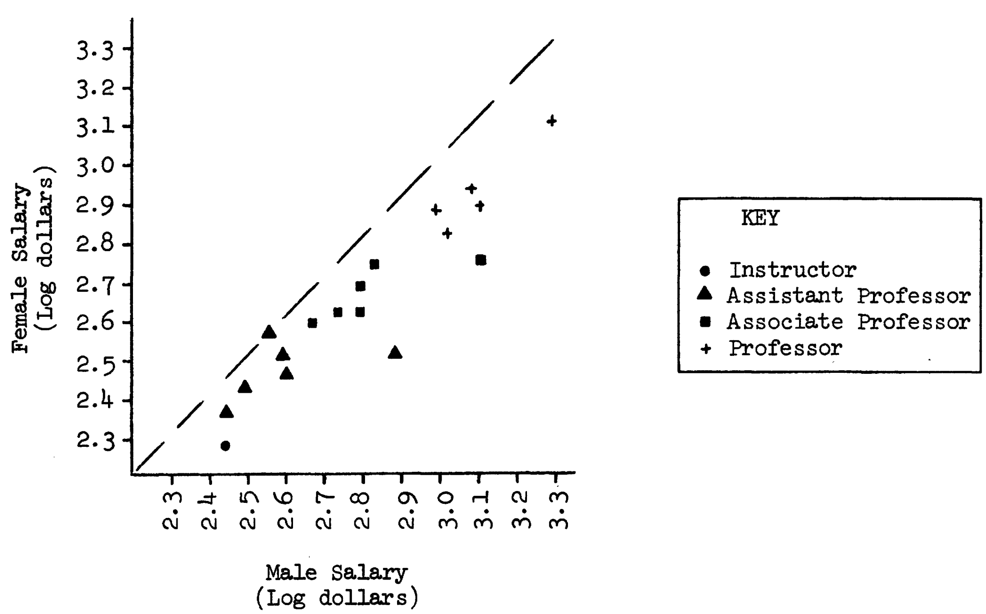


Figure 5. Number of Articles Published (# A) Plotted Against Log Salary (L $)

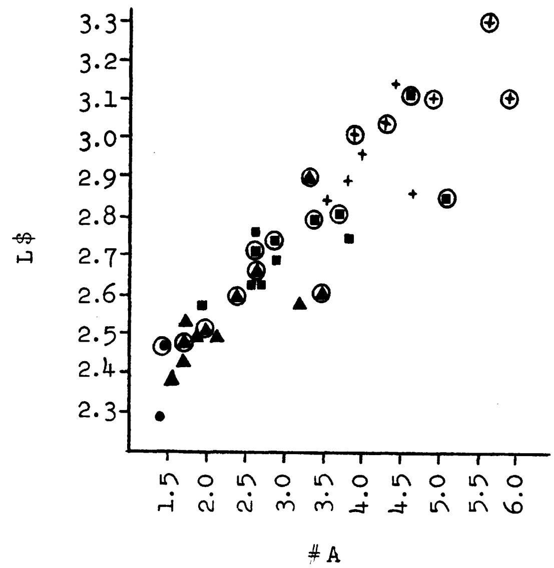


Figure 6. Number of Articles Published for the Male and Female Groups

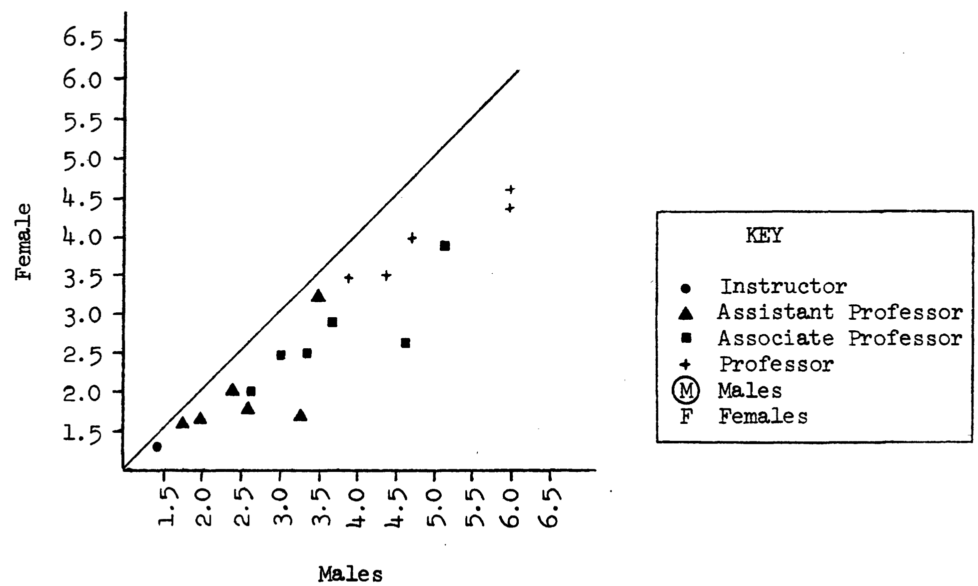


Figure 7. Number of Books Published (# B) Plotted Against Log Salary (L $)

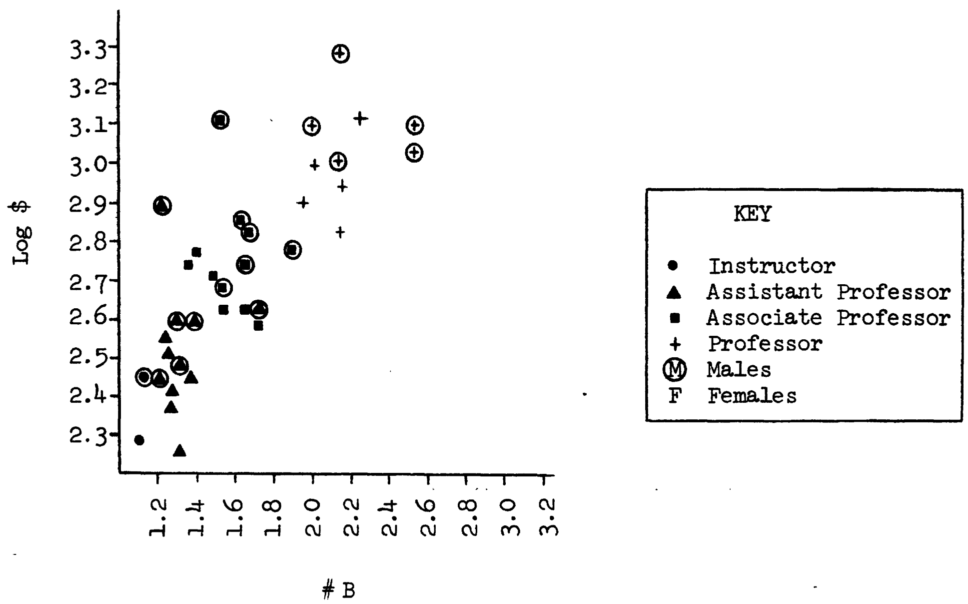


Table 1: Means and Standard Deviations of Faculty Salary (in units of $1000) by Academic Rank (R) and Departmental Affiliation (D)

		Dollars (in thousands)				Log Dollars			
		Females		Males		Females		Males	
R[1]	D[2]	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	1								
	2	9.77	2.77			2.23	.328		
	3								
	4								
	5								
	6	10.48	2.87			2.32	.243		
	7	9.51	1.80			2.24	.182		
	8	10.02	2.19	11.96	2.97	2.28	.207	2.45	.245
	9	11.00	3.55			2.36	.256		
2	1								
	2	12.02	2.72	13.76	2.50	2.46	.276	2.61	.180
	3	13.27	2.76	13.65	4.47	2.56	.210	2.58	.230
	4			13.30	2.37			2.57	.167
	5			14.32	1.87			2.65	.132
	6	12.49	2.15	13.61	2.50	2.51	.163	2.59	.173
	7	10.92	2.40	11.82	2.09	2.37	.211	2.45	.178
	8	11.59	2.40	12.30	2.07	2.42	.259	2.49	.156
	9	12.90	3.65	19.72	7.53	2.52	.245	2.89	.476
3	1			16.72	4.11			2.75	.454
	2	14.22	3.40	16.47	2.78	2.61	.338	2.79	.168
	3	15.77	3.44	17.27	3.49	2.74	.210	2.83	.194
	4			15.48	3.16			2.71	.279
	5			16.00	2.79			2.76	.168
	6	15.02	2.66	16.72	3.10	2.69	.180	2.80	.171
	7	13.74	4.24	14.71	3.01	2.58	.271	2.67	.194
	8	14.04	3.19	15.65	3.67	2.61	.308	2.73	.205
	9	16.29	5.38	23.72	8.62	2.75	.289	3.10	.371
4	1								
	2	18.39	3.93	20.15	4.14	2.89	.209	2.98	.221
	3	18.93	5.12	22.88	6.30	2.90	.266	3.09	.306
	4			21.97	4.90			3.07	.222
	5			22.51	4.69			3.09	.209
	6	19.20	3.61	22.56	5.14	2.94	.203	3.09	.226
	7			19.46	3.53			2.95	.183
	8	17.47	3.97	20.97	4.67	2.83	.262	3.02	.227
	9	23.44	6.14	27.35	7.49	3.12	.255	3.27	.272

[1] 1 - Instructor
2 - Asst. Professor
3 - Assoc. Professor
4 - Professor

[2] 1 - Business
2 - Education
3 - Biology
4 - Physical Sciences
6 - Social Science
7 - Fine Arts
8 - Humanities
9 - Health

Table 2: Number of Articles (#A) and Books (#B) Published for Males (M) and Females (F) by Academic Rank (R) and Departmental Affiliation (D)

R	D	#A F	#A M	#B F	#B M	N_F	N_M
1	1					5	5
	2	1.31		1.26		105	13
	3					15	5
	4					8	5
	5					0	6
	6	1.57		1.28		47	13
	7	1.40		1.14		35	4
	8	1.30	1.42	1.10	1.12	122	26
	9	1.36		1.10		69	2
2	1					12	22
	2	1.78	2.62	1.35	1.71	184	55
	3	3.19	3.50	1.22	1.35	59	46
	4		3.54		1.14	10	69
	6		2.66		1.29	2	41
	6	2.02	2.32	1.34	1.33	93	100
	7	1.60	1.74	1.29	1.21	63	38
	8	1.67	1.99	1.30	1.34	178	120
	9	1.71	3.28	1.23	1.16	107	25
3	1		3.88		1.72	8	25
	2	2.53	3.32	1.55	1.87	143	68
	3	3.88	5.16	1.35	1.66	62	56
	4		4.28		1.45	15	92
	5		3.68		1.37	2	57
	6	2.93	3.52	1.53	1.69	102	95
	7	1.91	2.67	1.71	1.56	34	45
	8	2.52	2.96	1.68	1.66	124	103
	9	2.63	4.64	1.41	1.52	51	25
4	1					7	15
	2	3.56	3.88	1.95	2.14	113	94
	3	4.63	5.83	2.00	2.01	41	92
	4		5.27		1.87	12	137
	5		4.94		1.90	1	83
	6	4.00	4.65	2.14	2.58	69	140
	7		3.02		1.98	21	46
	8	3.47	4.30	2.13	2.53	76	148
	9	4.44	5.65	2.24	2.15	25	34

WHY IS ASSESSMENT NECESSARY?

Lester R. Frankel, Audits & Surveys, Inc.

In 1974 the Subsection on Survey Research of the American Statistical Association prepared a research proposal to assess survey practices. This action was initiated because it was felt that certain trends were becoming apparent.

First, there was a feeling expressed at a 1973 A.S.A. Conference on Sampling of Human Populations that there was a growing reluctance of people to participate as respondents in sample surveys. Second, there appeared to be a critical attitude on the part of the public and the users of data with respect to the quality of many surveys.

These two factors could be regarded as consequences of the great growth of the sample survey that occurred during the previous 20 years. This approach has become an instrument of public concern. Government policy decisions, wage rates in certain industries, marketing and advertising actions on the part of business, and even the survival of a television series are to some extent dependent upon the results of these surveys. Sometimes, in everyday conversation a person in order to impute authority to a statement he has made will use the phrase "According to a recent survey..."

Sample surveys are conducted by many types of organizations including government agencies and professional groups such as found in universities, non-profit organizations, marketing and opinion research companies and in many other types of private companies. In these cases people who conduct the surveys tend to be qualified. Surveys are also conducted by persons who are untrained or inexperienced in this particular application of statistics. Questionnaires are found in hotel rooms, on theatre seats, on guarantee cards accompanying newly purchases appliances.

From the view points of both the generation of data and the dissemination of the results the sample survey has become an ubiquitous part of modern life.

The reaction of the public is of great concern to us in the profession of data collection. Without the confidence of the public and its cooperation in supplying basic data we would not have a viable mechanism. To a great extent its reaction is dependent upon what practitioners in the field of data collection are doing. Are we producing reliable information? Are we abusing respondents? Are we using sound methods?

These are the types of questions that the A.S.A. program to assess survey practices is designed to answer. Back in 1973 it was felt that sooner or later such information would be desirable in order to maintain a sound public image, to encourage sound research, to improve the quality of survey results and to eliminate some of the abuses. It was felt that in the distant future a handbook on standards of quality for surveys could be prepared.

It turned out that because of a series of events that started slowly in 1974 and mushroomed during the last year, the need for such an assessment is sooner rather than later. I would like to present a number of reasons why it is important that the assessment proceed as fast and thoroughly as possible.

. . . . . .

The first relates to the actions of the Federal Government. The United States government is the largest user and generator of survey data in the world. Techniques and procedures of survey sampling developed here tend to be emulated by all practitioners. The control of such operations is in the hands of Congress and in the executive branches.

What happened in the White House in 1973 called attention to the vast amount of records about individuals maintained by the Federal Government and the vulnerability of these records to unscrupulous persons. In 1974 three laws were enacted by Congress designed to minimize this possibility. Nevertheless they had the effect of making survey operations of the Federal Government more difficult.

These laws are (1) The Privacy Act of 1974, (2) The Freedom of Information Act of 1974 and (3) the Buckley Amendment to the General Education Provisions Act. While these acts were designed to protect the privacy of the individual they did not distinguish between statistical records and administrative records. In a statistical record the identity of the individual with respect to the observation about him is irrelevant. For an administrative record, however, the entire purpose of the record is to get information about the person as such. Had the enactors of this legislation understood and were aware of how data are actually being used for statistical purposes then perhaps these laws would not be as restrictive to survey operations as they now are.

There is some pending legislation at the federal level which would tend to impose additional restrictions on the making of surveys by and for the Federal Government. Among these are, first, the Wicker Bill to restrict the use of Internal Revenue Service lists to the agency itself and not to allow anyone outside to make use of these lists. This would make the taking of the Censuses of Retail and Wholesale Trade, and the Census of Manufacturers almost impossible. At the time when I prepared this paper there were a number of other bits of legislation pending. Among these is the Ashbrook Amendment as part of the House bill for a mid-decade census. This amendment would change the requirements for mandatory reporting to a voluntary system for all Censuses including the decennial population census. Since this provision is not in the Senate Bill, the Senate House Conference will determine its fate.

Another item of legislation is a House amendment to Section 13 of the Federal Energy Administration Act of 1974 that would add a provision prohibiting the Administrator from collecting or disseminating information on public opinions, attitudes or views either directly of by contract.

Regardless of the present status of these pieces of legislation it would appear that legislators are not aware of the operations of surveys in terms of the procedures used. The consequences of some hasty and seemingly minor action may have major repercussions. The assessment program of the A.S.A. will produce a document that will describe the practices that actually occur in survey operations. It should indicate that there is no motivation on the part of the survey statistician to invade anyone's personal privacy.

. . . . . .

Now, let us examine the need for assessment as it relates to the public relations aspect of our field.

During the past year the media have become very interested in the issues of confidentiality and privacy. Reports, commentators and editors have interpreted these issues as they relate to survey operations in the light of their knowledge of survey practices. These interpretations have a decided effect upon the attitudes of prospective respondents and as we shall see to some extent upon legislators.

I would like to review the publicity we received since the last annual meetings of the A.S.A. that were held in Atlanta. At these meetings a number of sessions dealt with privacy and confidentiality as they relate to statistical applications. A reporter of the New York Times, who was well versed in the objectives and techniques of survey taking was present. He took notes and spoke to many of the speakers and participants.

He prepared an article which appeared in the Sunday Times of October 26, 1975. The title was, "Polling Encounters Public Resistance; Decision Making Process is Threatened." The article dealt with many of the current problems involved in maintaining privacy and confidentiality, the apparent decline in response rates and the consequences of futher declines. It was a good article and treated the issues on an objective basis.

The story, sent over the New York Times News Service, was picked up by subscribing newspapers and columnists and appeared in many areas of the country. In this process the local editors were able to express their feelings and direct their influences since they had the options of making up their own headlines. First, let us examine some of these headlines:

1. Original article - Sunday Times, October 26, 1975.

2. Shreveport Journal - October 31, 1975. "Decline in Response Rates Worries Survey Experts"

3. Arkansas Gazette - November 2, 1975. "Twitching Curtain Phenomenon Makes Polling Harder"

4. Fort Lauderdale News - November 9, 1975. "Twitching Curtain Phenomenon - Oversaturation Threatens Future of Surveys"

What is this "twitching curtain phenomenon?" Reading from the original article,

> "It's called the 'twitching curtain phenomenon' by one survey expert. The interviewer approaches a suburban home clutching a questionnaire. Hamburgers are sizzling on the patio grill, the television flickers in the den, but the only sign of human habitation elicited by the visitor's persistent knock is a slight twitching of the curtain."

In an article unrelated to the Sunday Times story.

5. National Observer - November 1, 1975. "The Invisible Ink Caper"

The National Observer, a weekly newspaper, had commissioned a mail survey to determine the characteristics of its subscribers. The research company that was selected specialized in this type of operation. First, they selected a sample from the subscription list and then mailed out a questionnaire to each person in the sample under the name of the publication. The questionnaire bore the title, "A confidential survey among our subscribers." Every subscriber who was asked to fill out the questionnaire had good reason to believe that his individual response was, in fact, confidential. He was not asked to sign his name and in addition, there was a question which asked, "In what state do you live?'

One of the persons in the sample was a professor of optics at the University of Wisconsin. He was in his lab one evening and was about to fill out the questionnaire when, "for the heck of it" he put the form under an ultraviolet light--and a set of code numbers showed up. He was quite upset about it and at the suggestion of the director of the Wisconsin Survey Research Laboratory, wrote to the editor who published the story entitled, "The Invisible Ink Caper."

Now, for some of the comments.

6. Washington Evening Star - November 11, 1975. Editorial, "Trouble for Poll Takers"

I would like to quote part of this editorial. It referred to the story sent by the New York Times News Service and noted the decline in response rates. Then, "Some external factors are blamed--like fear of crime that makes householders reluctant to open their doors. But the more persuasive explanations have to do with the onerous practices of the surveyors themselves..."

"The surveyor who seeks indulgence to ask a few questions is likely to have an interminable list. Instead of a couple of moments, it takes many minutes to satisfy the statistical thirst of a client who may or may not have the public interest at heart. The aim and sponsorship of the survey frequently is kept obscure. The questions, after an innocent start, may get increasingly personal. If you escape queries about sex and birth-control practices, be prepared for the inevitable one about how much money you earn. When invasion of privacy is not the primary insult, there are seemingly dumb, ambiguous questions to cope with. Is someone playing a psychological game, probing in devious ways to locate hidden attitudes? The surveyor, in one variation, may turn out to be leading up to a sales pitch for some unwanted product."

"Pollsters bemoan the resistance of survey targets on public-service grounds, voicing fear that government and business will have trouble assessing our needs and desires. But what did our past patience with supposedly scientific sampling get us? Inane television programs bracketed by inane commercials. Junk on the supermarket shelves. Political candidates that are all image and no substance. Large, inefficient automobiles that supposedly were the only kind marketable in this country."

"It will be a small loss if the surveyors eventually find every door closed."

7. Indianapolis Star - December 4, 1975. Jim Flebig, columnist, "Mums the Word to Stop Pollsters"

I have cited these news items to illustrate and to indicate the types of thinking that are in the minds of people who can very easily influence public opinion. When someone reads these articles and editorials he or she may get the impression that the practices so elegantly described are the rule. How often does the twitching curtain phenomenon occur?

Are all mail questionnaires serialized and coded with invisible ink? What about the charges brought forth in the Washington Evening Star? The Evening Star, being published in Washington, is read by many Congressmen. To what extent had they been influenced when they vote for legislation making responses to the Census questionnaire voluntary rather than mandatory.

In order to respond to charges such as these we ourselves have to know what goes on in our field. With the exception of some major government and non-governmental studies we really do not know. The assessment program should fill this void.

. . . . . .

One of the long-range programs of the assessment program is to set up guidelines and standards for insuring the quality of survey research data.

The general public and segments within are being asked to rely upon the validity and accuracy of surveys. We are all aware of the uses made by government administrators of the data generated by continuous data collection surveys. The procedures have been open to public examination and scrutiny by scholars, review boards and by the Statistical Policy Division of the Office of Management of Budget. In the case of ad hoc surveys conducted by and for the executive branches, OMB clearance is required and as a result certain standards are met.

However, for the most part surveys are conducted where no clearances are required and where no quality standards exist. At the present time there is a great emphasis upon comparative advertising--where one brand is compared against another. Surveys are conducted to make claims for certain advertisers; other surveys are made to counteract the claims of competitors. The FTC comes up with another survey to show that consumers have been misled. Even though the FTC is part of the Federal Government, its surveys do not require OMB clearance because it is a regulatory agency. When dissatisfaction with the survey findings occur, surveys are challenged.

Because of the lack of quality standards, when there are disputes, legal steps are taken. Statisticians are engaged by both sides. Adversary conditions are created. Lawyers are given cram courses in sampling so that they can be in a position to challenge an opposing statistician's interpretation of the laws of probability. If standards were established for the conduct of sound research, then much of the bickering would be eliminated. Challenges would be made on the basis of relevancy and basic definitions and not upon the measurement techniques.

. . . . . .

In carrying out the assessment program it can be expected that many benefits of a technical nature will occur. Over the past 40 years statisticians have worked to improve the survey instrument. The approach has been to examine each of the component parts of a survey and attempt to increase the efficiency of each of the parts.

The problems of sampling, interviewing, questionnaire design, refusals, not-at-homes, field travel, data processing, and estimation were all treated separately or at most in pairs. The concepts of the mean square error and amount of information per dollar were about as far as we advanced in coordinating the parts of a survey.

When the assessment program gets underway it will become possible to look at a survey as a system. The assessment team will examine a large number of different surveys each with its own characteristics. In the analysis it will be possible to see how the different parts of the survey interact--how variations in one component affect the other components--and their effect on the overall accuracy.

. . . . . .

A final point that I would like to discuss is why should the American Statistical Association undertake such a program? My answer is simply that if the A.S.A. does not undertake this step it will be done as a by-product of some other operation and may be done by people who are inexperienced in survey methodology.

The persons who formulated the pieces of legislation of 1974 that I referred to previously did not distinguish between statistical or survey data and administrative data.

During the past year a number of groups have begun working in the area of confidentiality, privacy and the right of human beings.

One group is particularly interested in survey operations in this connection. This is the Panel on Privacy and Confidentiality as Factors in Survey Response under the guidance of the Committee on National Statistics of the National Academy of Science -- National Research Council. It is obtaining empirical evidence to determine whether and how much the public's concern about privacy and confidentiality has an effect upon the response and accuracy of people selected for sample surveys. The members of this panel have had vast personal experience in their own fields. However an idea of survey practices on a broader base is necessary to enable the panel to perform its task.

There is still an official committee involved. This is the President's Committee on Privacy and Confidentiality. It is studying all aspects of these issues and is to come up with recommendations for legislative action. When the group was first formed it had not considered the role of surveys. Yet much of the legislation that they will recommend will have a direct effect on the future operations of sample surveys. The situation has now been corrected and the consequences upon survey operations are now being considered.

The third group I would like to call attention to is the National Commission for the Protection of Biomedical and Behavioral Research which was created to develop guidelines for Institutional Review Boards in judging research proposals. The Commission asked Professor Donald T. Campbell to prepare a background paper covering Evaluation Research and Social Experimentation. Campbell felt that to cover the topic he had to include and make recommendations to cover Social Indicators, Secondary Evaluations and "to some extent Survey Research." Professor Campbell sent an early draft of his recommendations to people in these five fields requesting comments. The comments he received, I am sure, reflected the personal experiences of the people who wrote him. If the assessment program had been under way a sounder set of recommendations concerning survey research would have evolved.

PROGRESS AND PROBLEMS IN THE ASSESSMENT OF SURVEY PRACTICES

Barbara A. Bailar, U.S. Bureau of the Census

I. Description of Project Proposal

In addition to the growing reluctance on the part of the public to participate in surveys, there is a growing critical attitude on the part of the public about the quality of the surveys conducted. It is not unusual for two surveys on the same topic to reach different conclusions. These differences may be caused by a difference in questions asked, a difference in the sampled populations, a difference in the survey methods used, a difference in the controls of collecting and processing the data, or other such factors.

In order for the survey community to develop effective programs to correct these problems there must be some way of assessing survey practices, particularly as they affect the quality of the data gathered. Such an assessment would provide guidance on the contributions of survey research as well as the limitations and abuses of the method. This assessment would also provide a defense of the use of surveys to those who would like to see them eliminated altogether.

This is an excerpt from the proposal, subsequently funded, which the ASA Subsection on Survey Research Methods submitted to the National Science Foundation. The Subsection realized that the most crucial as well as the most difficult part of assessing survey practices would be in the development of criteria for describing and classifying specific aspects of survey methodology. For this reason the proposal was for a developmental and feasibility study of the methods and practices of survey research, rather than a nationwide study of such practices.

In the proposal the long-term aims of a full-scale study are given as:

A. the development of a set of specifications by which survey research practices can be assessed,
B. the determination of the numbers and kinds of surveys being conducted,
C. the preparation of a profile of survey practices and the state of survey methodology as it is now practiced with emphasis on problems that vary in incidence between government, private, and academic research, and
D. the encouragement of the working together of representatives of various survey groups to improve the quality of survey results.

Because the Subsection was not clear before doing a good bit of developmental work that those aims could be met, the aims of the developmental and feasibility study are more limited. The aims of the feasibility study are:

A. to develop sampling frames for surveys sponsored by the Federal government, by state and local governments, by private non-commercial survey research groups, and by private commercial survey organizations,
B. to develop a list of information items and the specifications of the criteria by which surveys can be evaluated, and
C. to conduct a pretest of the general approach.

A product of the feasibility study will be a report including a description of the sampling frames, data items collected, response classifications developed, and problems encountered. Detailed and comprehensive recommendations for the full-scale survey will be presented.

This, then, is a description of the project and what is to be accomplished. Work is in progress and we are now convinced, from the results of the ongoing pretest, that the project is not only feasible but critical.

II. Definition of Universe

To accomplish the aims of both the developmental study and the nationwide study, individual surveys are the subjects of interest. All of the surveys that are conducted by the Federal government, those conducted by State and local governments, those conducted by non-commercial survey research groups, and those conducted by private commercial survey organizations are in the universe of interest. However, since the description of survey practices should relate to survey research as it is currently practiced, the universe is limited to those surveys funded or carried out in 1975. For the purposes of the developmental study only the universe is further limited to those surveys that pertain to human populations.

III. Development of Frames

There are alternative ways of constructing frames for the selection of a sample of surveys from each of the four sectors. One might consider compiling a list of all survey sponsors and then making a list of the surveys for each sponsor. Or, alternatively, one might consider making a list of all survey organizations or survey takers and making a list of the surveys conducted by each survey taker. Either of these approaches would be satisfactory. However, one approach is easier than the other for certain sectors but not for others. Also, there is no complete listing of either type. Currently, we are constructing multiple frames of each type. These, of course, introduce the accompanying problems of multiple probabilities of selection for a survey. The problems of developing frames vary from sector to sector. Let us consider these problems one by one, first from the point of view of developing a list of survey sponsors and then by developing a list of survey takers.

A. Sponsors of Surveys

1. The Federal Government

Where would you go if you wanted a list of surveys sponsored by the Federal government for a specific time period? Perhaps you think first of the Office of Management and Budget (OMB), which, under the provisions of the Federal Reports Act, has the function of clearing forms for some agencies and departments for which ten or more people are asked to provide information. These forms include applications, such as passport applications, program evaluation forms, statistical surveys, management forms, and a miscellaneous group. However, OMB does not clear forms for the regulatory agenices of the Federal government; the General Accounting Office (GAO) does.

With this information in mind, it was our intent to ask OMB and GAO for a list of surveys completed in 1975. It quickly turned out that this information is not available. OMB and GAO have lists which show the clearance date of a form and the expiration date. OMB has a quarterly listing that shows the names and numbers of all forms that have not expired as of the end of the quarter. Since some of these forms could have been cleared 3 years earlier and some 1 month earlier, this list represents a mixture of time periods.

Another problem that arises is what forms should be included as surveys. The only type that can always be excluded is applications. All of the other groups contain at least some forms that could be classified as surveys. There is also some confusion about whether forms represented surveys or not. To decide this, one had to look at the individual dockets and decisions had to be made about what would be included. Finally, a list was compiled of about 450 forms which had been cleared in 1975 that were thought to represent surveys of human populations.

However, the clearance numbers do not uniquely identify a survey of interest. One survey of interest has two clearance numbers. The Current Population Survey (CPS) and its control form have two different clearance numbers. One would not want to sample the control card only. So there are some problems of putting several numbers together as a single sampling unit.

The GAO cleared about 200 forms in 1975 and made these lists available. Most of those that were surveys were surveys of businesses not people, so all but four were excluded from this study.

From these two sources now all the surveys commissioned by the executive branch of the Federal government in 1975 either to be carried out by an agency or by a contractor were represented. However, as was found later, the clearance of a form did not guarantee that a survey was carried out.

Yet surveys carried out as part of a grant which may also be federally funded are usually not cleared by the Federal government and so do not appear in either the OMB or GAO files.

There are two strategies that we are currently investigating to get a list of surveys completed under grants. The first strategy is to contact the Smithsonian Science Information Exchange, Inc. The Exchange has a computerized system of information about research conducted under sponsorship by the Federal government and foundations and other institutions which fund research. However, listings in the Exchange are voluntary. They appear to stem from a large number of sources, the nature of which is not known.

Recently, we requested some information from the Exchange on some special topics, just to get an idea of the amount of survey research that is covered by grants. Of 44 research studies on the general topic of racial attitudes and prejudices, no summary of the project was given for four studies, but of the remaining 40 studies, 23 involved some use of surveys.

There are two problems with using the Exchange as a source of surveys. First, the listings are not complete, since it is voluntary. Second, the topics which might conceivably be areas in which surveys are conducted are almost inexhaustible. Therefore, the topics themselves would constitute an additional stage of sampling at best.

A second approach to grants is to go directly to the granting source. This means developing a list of granting organizations. It is possible, though not easy, to do this for Federal agencies.

Another kind of survey sponsored by the Federal government but not included in those already mentioned is a survey funded by either the legislative or judicial branches of the government. One of these which recently was reported in the papers was a survey on gambling, sponsored by the Commission on the Review of National Policy toward Gambling, a joint congressional commission. At the present time a survey such as this one could only be represented on a list provided by a survey organization, since no list of sponsors includes congressional committees. It may be that we shall need to develop a list of possible sponsors for each of the legislative and judicial branches of government.

2. State Governments

To develop a list of surveys sponsored by the State governments, one needs to contact each State government. In some states, it is the State Planning Director and in others the State Budget Officer who is most familiar with survey plans. Letters have been sent to the planning directors and budget officers of each of the 50 States and and the District of Columbia asking them about whether or not there is a central clearance agency for statistical forms and, if not, whether or not they will provide a list of State agencies to be queried directly. Replies from 37 of the States have been received. One State refused to provide any information. All of the others have said there is no central clearance agency. One possible exception is Hawaii which by law requires that copies of all state studies be lodged in the archives. Each of the other States has provided lists of agencies. Many have asked to be kept informed of what we find out and have shown a lively interest in the study.

With a fair amount of effort and considerable letter-writing a list of surveys sponsored by each State, except the one which would not cooperate, should be compiled.

3. Local Governments

There are so many local governments that it would be virtually impossible to contact all of them. For example, there are city governments, county governments, school board districts, and other units of local government that could sponsor surveys. The National League of Cities has a publication called The Mayors of America's Principal Cities, which gives a focal point to make inquiries. However, we have been advised by the Director for Policy Development and Analysis of the National League of Cities and by the Director of the Data Services Center of the International City Management Association to limit attention to cities of about 25,000 or more. They feel that smaller cities would not have the resources to sponsor surveys.

No contacts have yet been made with local governments. In one State a list is being compiled of all the units of local government. Until that list is available, no further steps will be taken.

4. Universities and other non-profit organizations

Departments in universities may carry out surveys for use by the university, by the department faculty, or by the students. Many of these surveys are small-scale efforts. We plan to inquire about surveys sponsored by departments at two or three of the larger universities to get an idea of how much survey work is going on in this area.

Many universities also have survey research centers or other types of non-profit research facilities that may sponsor and carry out surveys. A listing of the survey research centers was available from the Survey Research Laboratory at the University of Illinois.

The Research Centers Directory published by the Gale Research Co. includes research institutes, centers, foundations, laboratories, bureaus, experiment stations and similar nonprofit research facilities. To be listed in this directory, an organization must have two key characteristics: (1) it must be formally identified by a specific or distinctive name or title, and (2) it must be established on a permanent basis as a separate entity for carrying on a continuing research program. Therefore, one wouldn't find a listing for a specific department at some university.

The listings in the directory are under 16 main headings, with an addendum and also periodic supplements. Approximately 5,900 listings are included for 1975 with a brief summary of the type of research being carried out. Of the 16 main areas, three of them seemed to have no organizations that might carry out surveys. They were astronomy, physical and earth sciences, and regional and area studies. Of the remainder, some seemed to carry out surveys. Over 500 letters were mailed out to those which seemed likely to be doing surveys. About 70 percent have responded to the inquiry, some in great depth. Many, many of the organizations are carrying out a very small number of surveys.

In a nationwide study, more listings from this directory would have to be contacted. For example, there are 949 listings in the "life sciences" area. Letters were sent to only 48. We had hoped to get a definitive reading on whether hospitals, biomedical labs, psychiatric institutes, and others of this type conducted surveys. Of the 48 places contacted, one is no longer in existence, and 36 have answered the request. About half of them do surveys.

Unfortunately, many of the research groups not only sponsor surveys but are also takers of surveys. So this group of listings is a mixture of sponsors and takers. Also, many of the listings in the Research Centers Directory are not academic centers. Thus, they would represent the non-profit-making sector.

In addition, in the non-profit sector are the large foundations. Reports are available from these foundations that give a list of topics for which grants were given in 1975. It is possible to compile a list of studies which may have involved survey work. The Foundation Directory and The Foundation Grants Index published by Columbia University Press are the sources of the list of foundations.

5. Private Commercial Sector

To get a list of all surveys sponsored by the private commercial sector means getting a list from all private companies of market studies they have sponsored, from political sources of polls or studies they have commissioned, and from newspapers and magazines of the surveys they have instituted.

It has been suggested to us by persons in the private sector that the best way to get a listing of market research studies is by contacting the large companies who sponsor such studies. A list such as the Fortune "500" largest corporations would doubtless provide the largest share of the commercial survey research. This may be the only way to find out about many studies because, in many instances, the organization that carried out the study cannot reveal the name of the client. However, there will be gaps in the frame because of the omission of small companies that commission surveys.

There is a publication Daily and Weekly Newspaper List and Magazine List published by Luce Press Clippings, Inc. that would provide a list of newspapers and magazines to be queried about the surveys they have sponsored.

It does not seem feasible to contact politicians themselves to ask about polls they may have commissioned. Except for individual polls for particular candidates, the political polls would be represented in the list provided by survey takers.

B. Survey Organizations or Survey-Takers

An alternative way of compiling a list of surveys is to go to the survey-takers. For some sectors this is an easier method of finding out about surveys.

1. Federal Government

Some government agencies carry out many surveys - the Bureau of the Census, the Bureau of Labor Statistics, parts of the Department of Agriculture, etc. However, all of these surveys would be represented in the list of surveys cleared by OMB or GAO. A few agencies such as the Internal Revenue Service, the Central Intelligence Agency, and others which are exceptions to the Federal Reports Act might be queried about survey-taking. However, for the executive branch, the list of survey sponsors seems to be the preferred method of constructing a frame.

2. State Governments

By and large State governments are not takers of surveys. They may do a few mail surveys. However, the surveys that are actually conducted by the State would be revealed in the list of sponsors. Many State governments commission the State Universities to carry out surveys. These surveys would become known to us from the university sources.

3. Local Governments

If a local government carries out its own survey this will be made known to us when the local unit provides the sponsorship of the survey. In some metropolitan areas, a council of governments or municipal leagues may carry out surveys. A listing of these organizations may give us an additional list of surveys.

4. Universities and other non-profit organizations

The list of research organizations which is being compiled contains many survey-takers. These survey-takers can provide information on the sponsorship of the surveys they have undertaken. The Federal surveys of the executive branch would be covered elsewhere. But a number of the surveys conducted are for associations or groups that would not be represented on the sponsor list. Also, the surveys sponsored by the legislative or judicial branches would not be represented elsewhere. Therefore, this list provides surveys that would not be covered by the sponsor list.

5. Private Commercial Sector
Many of the surveys conducted by the private organizations are for the executive branch of government and would be represented on the sponsor list. But many would be for industrial concerns. At present we are compiling a list of private commercial organizations that carry out surveys. This may be an alternative way to include the market research studies. However, in this case, the organization would have to contact the client before we could even know the name of the client for which a study was done.

To compile a list of survey organizations, many sources of information have been suggested and used. The American Association for Public Opinion Research (AAPOR) published a list in January, 1975 of 123 agencies and organizations represented by their membership. The organizations are of both the commercial and non-commercial types. The American Marketing Association published a 1975 directory of both marketing services and members. This directory added many organizations to the previous list. At the end of the directory is a vocational listing by universities, colleges, and schools, and by private companies. This vocational directory pulls in companies that are not in the market research business but are large companies that often carry out surveys. Another source of information was Bradford's Survey and Directory of Marketing Research Agencies in the U.S. and the World.

Using these sources, a list of several hundred organizations was made, yet it is still incomplete. A perusal of the yellow pages of the Washington telephone directory turned up 23 different headings under which organizations that carry out surveys could be listed. Most of these organizations, except the very largest, were not represented in the earlier list. Thus, it seems clear that to develop a frame for these organizations, the yellow pages of telephone directories, at least for certain large cities, must be used.

IV. Description of Pretest

We have developed a questionnaire that has 15 separate parts, not all of which have to be explored for every survey. These 15 sections are:

1. Research problem
2. Responsibility for survey
3. Questionnaire design
4. Sampling design
5. Data-gathering activity
6. Mail questionnaire surveys
7. Personal interview surveys
8. Telephone interviews
9. Data-collection problems
10 Coding procedures
11 Keypunching
12 Machine editing
13 Tabulation
14 Report writing
15 Dispostion of data

To find out whether this questionnaire would give us the kinds and types of information necessary to assess the quality of the survey, a pretest was necessary. Since the OMB list of Federally sponsored surveys was available and easy to work with, we selected 25 surveys from that list to use in the pretest. The surveys were not a random sample but a purposive selection designed to include different kinds of surveys, problems, and types of contractors. Surveys were classified into two groups--methodological and subject-matter oriented. They were also classified by the method of data-collection--personal visit, telephone, or mail. They were also classified by whether the survey was conducted by the government agency, by a university or non-profit organization, or by a commercial organization. Surveys of all of these types were included. At the present time the field work for most of these surveys has been completed.

To this list of 25 Federal surveys from the OMB list, we plan to add two from the GAO list, five from the State or local governments, and at least three other surveys. This last group will represent surveys sponsored by organizations not appearing on a sponsor list, political polls, and newspaper surveys.

There are usually at least two interviews for each survey - one with the sponsor of the survey and one with the contractor. In the cases in which the sponsor and contractor are the same, separate interviews with groups responsible for different aspects of the survey are usually necessary. Each interview takes no less than 1 hour and usually runs about 1 1/2 hours. The sponsors of the survey can provide information on the objectives of the survey, how they selected a contractor, the cost, the questionnaire, and sometimes more. The contractor provides information on the sampling design, the field work, response rates, coding, and other such problems. In some cases, a subcontractor provides needed information.

V. Results of Pretest

Initially, the effort to assess surveys was thought of only in terms of specific survey operations. However, the Steering Committee for this project decided that some attention should be paid to the question of whether a survey met its stated objectives. Therefore, the assessment falls into two categories:

Did the survey meet the stated objectives?

How well technically was the survey carried out?

A. Accomplishment of Objectives

The objectives we are considering are the subject-matter goals, the objectives that were stated as the reason for carrying out a survey at all. We are not considering the survey specifications, which could be met without satisfying the survey objectives.

Determining whether a survey met its stated objectives is a somewhat subjective judgment. However there are specific cases in which a clear-cut decision can be made. Some of these are as follows:

1. The design of the survey does not permit the survey to reach its objectives.

   An example of a survey of this type is one in which the objectives are specified that survey data will be used to predict the performance of a group of people who have been exposed to a certain type of educational experience. It is desired to be able to identify factors associated with success. However, the design of the study calls for the people to be studied within a very short time of their exposure. It is probably too soon for them to have yet been successful or unsuccessful.

2. The results of the study are never made available in written form and thus are never disseminated.

   The purpose of a survey is usually to learn something about a population, a program, or a methodology. The results may never be written up so that it is hard to know whether the objectives were met.

3. The design specifications contain either conflicting or inadequate detail.

   Many of the surveys commissioned by the Federal government are carried out by contractors. The usual method is for the agency to put out a Request For Proposal (RFP) and invite survey organizations to submit bids. Many contractors find conflicting or inadequate specifications. If a counter-proposal is submitted, the contractor may not be granted the work. Many RFP's specify inadequate survey design.

B. Technical Conduct of Survey

It is now apparent that there are certain areas in the survey process which present difficulties. This presentation will focus on only a few major areas.

1. Calculation of nonresponse

   There are two separate kinds of surveys to be discussed--those that use random digit dialing and all others. Let us concentrate first on those which do not use random digit dialing.

   Many project officers do not understand the hazards of a low response rate. They often specify as acceptable low response rates in the RFP. Many contractors do not see it as their function to enlighten project officers about response rates. Some project officers are not aware of and are not interested in the size of the nonrespondent group. Some do not understand some of the adjustment that is made for nonresponse and thus do not know what the nonresponse rate is. Some of them have specified an acceptable level of response in their research proposal; and if the contractor presents a report which indicates that level, all appears well.

   A few illustrations may clarify these problems. One project officer told us that the response rate was over 90 percent. This high response rate was because of the foresight of the contractor who arranged that each cell would have back-up samples available. The true response rate was about 56 percent.

   In another case the response rate was not yet at the specified level of 70 percent. The contractor had certain interviewers work until the level was reached. In a case such as this, there is probably a very big difference between respondents and nonrespondents, especially in the hours at which they are home.

   Random digit dialing has brought a new dimension to problems in calculating nonresponse rates. By some process, telephone interviewers are furnished with a set of telephone numbers to call. Some of these numbers do not answer or are busy. A nonresponse rate is often reported in which the denominator is only the number of calls answered. Rarely is the number of "working banks" of numbers within specific exchanges known. Thus, a response rate of 80 percent may be reported for a case in which 20 percent of the numbers were never reached. In many cases, this neglect of keeping adequate records of the results of each call results in quota samples. Thus, in one survey a quota of a certain number of completed calls was assigned to each region for which estimates were to be made, and no nonresponse rates were calculated at all.

2. Representativeness of the sample
This problem is closely related to the problem of nonresponse, but has some additional features. For example, a sponsor may say that his objective is to have estimates for the entire United States. If he selected random digit dialing, he is limiting the sample to telephone households, which have different characteristics from households without telephones. In one survey, blacks were excluded because it was too hard to get enough of them to make good estimates for blacks alone. So, on a topic for which one might expect a difference in behavior between blacks and whites, no information was provided.

Nonresponse, of course, is a serious problem in the representativeness of the sample data. Some of the time the problem is addressed in the tabulation of the data, where weights are applied or changed in an effort to make the data more representative. This is not done routinely, though. Even when it is done, the weights merely inflate the known sample to represent those never contacted.

3. Recall problems
There seems to be little or no concern by the sponsor that many of the questions asked may be difficult if not impossible for a respondent to answer. A yearly recall on expenditures for certain items is not unusual. Recall over a period of years for certain types of behavior is common. Though there may be no alternative, very little if any research is going on to help in the determination of realistic recall periods.

4. Computation of variances
Many project officers do not use variances, even if they are produced. The technique used in computing variances seldom finds its way into published reports of the study. Frequently, though weights are used in estimating means, totals, and proportions, variances are computed using unweighted data. The usual method of calculating variances is to use the formula for simple random sampling no matter what the structure of the sample. To be fair, certain sponsors do recognize what is being done and are satisfied that their variance estimates are probably conservative.

5. Cost
Technically, cost is not a problem in carrying out a survey. It is included here because there seems to be no way of predicting the cost of a survey given the sample size. In the surveys we have studied, costs ranged from $5 a case to well over $300 a case. Differences in cost were not just functions of the method of interview. Thus, one mail survey with some personal interview cost about $26 a case while another that was mail with telephone followup was $80 a case. In both cases, the cost included processing and cleaning the data, as well as providing a tape. A personal interview survey cost $27 a case for one survey and $204 a case for another, in both of which the contractor was responsible for the processing and cleaning of data.

IV. Future Work

There are several things that must be completed for the feasibility project. The first of these is to complete the development of the frames.

A second area of concern is that the feasibility at this time has been tested only with government-funded surveys, and even those do not represent surveys conducted under a grant. We need to find out the following things:

1. Are there substantive difference between surveys other than government surveys conducted by the private commercial firms, so that the questionnaire developed for government surveys is not useful for the private sector?
2. Is the turnaround time of commercial surveys such as market research studies such that the data are too far in the past to be recovered?
3. Are the record-keeping practices of commercial firms such that the data would be available?
4. Since the commercial firms have no public obligation for reporting and

indeed may be bound to protect their clients' interest will they cooperate in such a study?

Because staffing incurred a late start in working on this project, we cannot do any testing in the private sector, except, of course, for those firms who do work under contract for the Federal government. We have received excellent cooperation from those we have visited, both commercial and non-profit organizations.

The final step in our work is to write a complete report on the feasibility study. This will take place in the fall and winter and we hope to give the National Science Foundation a finished report on this pilot study in the spring of 1977. We intend also to submit a project proposal for a full nationwide study.

ASSESSMENT OF SURVEY PRACTICES AND DATA QUALITY -- COMMENTS CONCERNING STATISTICAL POLICY

Joseph W. Duncan, Office of Management & Budget

The Statistical Policy Division (SPD) of the Office of Management and Budget (OMB) is vitally concerned with today's topic of the assessment of survey practices, especially for Federally-sponsored surveys. We have two specific involvements:

1. In our role of planning and coordinating Federal statistics, we are concerned with the quality of Federal statistical data and with using Federal dollars in a cost-effective manner to produce statistics that evaluate the social and economic well-being of the country. A large proportion of the statistical budget is spent on statistical surveys.

2. The SPD also has the responsibility for implementing the Federal Reports Act. This reponsibility involves clearing any Federal data collection effort that contacts 10 respondents or more. In carrying out this responsibility, we review all proposed statistical projects before the data are collected. It should be noted, of course, that the review of statistical proposals before they are carried out does not include monitoring of the actual conduct of the statistical surveys; only in the case of repetitive surveys do we review what happened in the previous wave before approving the recurring effort.

The paper on "Progress and Problem in the Assessment of Survey Practices" that Barbara Bailar presented reports on the feasibility project she is directing. The objective of this project is "assessing survey practices, particularly as they affect the quality of the data gathered." This project was funded by a grant from the National Science Foundation (NSF) to the American Statistical Association's (ASA) subsection on Survey Research Methods. Her report focuses on two aspects: (1) the developing of sampling frames for surveys of human populations and (2) assessing survey practices by actually conducting an interview with the survey sponsor and/or executor to obtain a detailed report for evaluating the quality of the survey. The development of the frame covers surveys sponsored by Federal, State, and local governments; universities and other nonprofit organizations; and private commercial survey organizations. The actual feasibility study to assess the quality of surveys was carried out using mainly those surveys sponsored by the Federal Government. The preliminary results presented here consider it possible to assess Federally-sponsored surveys.

Because of the interest of the SPD in the quality of Federally-sponsored surveys, the results of this study can prove invaluable to our responsibility vis-a-vis surveys sponsored by the Federal Government. To the extent that the results of this study provide specific guidance on how to evaluate the quality of Federally-sponsored surveys, we are prepared to act quickly in implementing the findings of this study.

Les Frankel, in his paper on "Why is Assessment Necessary?" describes the need for assessment of surveys from a broader point of view. In fact, the scope of the ASA's feasibility survey covers a much broader base; that is, <u>all</u> surveys of human population. Moreover, Barbara Bailar's conclusion from this preliminary study is that assessing the quality of surveys of human populations is feasible. Therefore, we concur with her and will stimulate the subsection on Survey Research Methods to propose a nationwide study to evaluate the quality of surveys.

This assessment project is important. In addition to the reasons cited by Bailar and Frankel and in the introduction to this discussion, there are several other important statements of concern. These include the Joint Ad Hoc Committee on Government Statistics and the Commission on Federal Paperwork.

<u>Recommendations of Joint Ad Hoc Committee on Government Statistics</u>

The <u>Joint Ad Hoc Committee on Government Statistics</u> (comprised of the American Sociological Association, the American Statistical Association, the Federal Statistics Users' Conference, the National Association of Business Economists, and the Population Association) has recently prepared a report which includes recommendations on Federal contracts for statistical services. The recommendation states:

> "<u>The Committee recommends a thoroughgoing review of the procedures for awarding Federal contracts for statistical services to nongovernmental organizations. The review should include after-the-fact audits of the quality of the services provided, performance in relation to time and costs, and a look into the comparative quality and costs of doing the work inhouse.</u>"

The report points out the following specific problems:

> "1. The Government agency may have little or no inhouse capability and, thus, cannot do an adequate job of developing specifications, evaluating bids, and supervising the execution of the contract.
>
> "2. There are no adequate standards for qualifying some contractors and disqualifying others in the light of their capability to produce survey or study results which meet acceptable standards of statistical quality.

"3. The practice of favoring the low bidder may be counterproductive, especially if there is no adequate inhouse capability to evaluate the quality of work which the contractor provides.

"4. Contractors may lack statistical staffs themselves and therefore subcontract an essential part of the project, thus further removing control over the execution of the project.

"5. Well qualified contractors may have so many projects going at one time that top personnel cannot provide adequate attention to a particular one and thus the work is, in fact, assigned to relatively inexperienced personnel.

"6. The contractor may not be in a position to develop the policy implications of a project. The agency staff may lack the statistical skills needed for a full evaluation and thereby be inadequately equipped to make the most effective use of the product."

Another recommendation of this committee on need for analysis includes the following statement:

> "Throughout the Federal statistical system, there is also a need for more resources devoted to the development and application of statistical methodology in the collection, analysis, and presentation of statistical data."

The discussion of this recommendation suggests that SPD should followup on nonissuance of results of data collection projects:

> "The collection of statistical data which does not result in tabulation, analysis, and publication is obviously a waste of resources. Clearly, at the time of clearance, SPD should be informed about plans for tabulation and issuance of the results. SPD should have a followup procedure to determine if and why such plans are not carried out in those instances in which this occurs."

Concerns of Federal Paperwork Commission

The Federal Paperwork Commission (FPC) proposes to undertake a study on survey methods used by commercial firms working for the Federal Government. They would like to investigate steps which might improve survey practices. At the same time, they would like to determine how such practices might be changed to reduce paperwork and respondent burden.

This study would focus on the development of the Request for Proposal (RFP), the evaluation of submitted proposals and letting contracts, and on monitoring of statistical work conducted under contract.

The need for an indepth study of contracting statistical work is obvious. The feasibility study which we are discussing today will begin to address this problem, although the objective of the ASA study is much broader.

Current Activities

While these investigations are being made, some recent actions at OMB are important as immediate steps to improve the quality of government-funded surveys.

The President's Reports Reduction Program. Guidelines for reducing public reporting to Federal agencies were issued by the President on March 1, 1976. For statistical surveys or reports, the following statement on response rates was given:

> "It is expected that data collections for statistical purposes will have a response rate of 75 percent. Proposed data collections having an expected response rate of less than 75 percent require a special justification. Statistical data collection activities having a response rate of under 50 percent should be terminated. Proposed statistical data collection activities having an expected response rate of less than 50 percent will be disapproved."

The purpose of disapproving statistical data collection activities with response rates below 50 percent is the expected poor quality of such data. The need to compute response rates correctly is highlighted in the paper prepared by Barbara Bailar.

Since the clearance process in the SPD looks at expected response rates rather than actual response rates, we do not always obtain an adequate report of the response rate actually obtained. In repetitive surveys we often request historical information before clearing a subsequent wave of the survey. The need for an adequate review of results obtained in statistical data collection efforts may determine a modification of our requirements to agencies in order to be able to improve our capacity to monitor the quality of Federal statistical surveys.

Estimate of Contractual Workload in Federal Surveys. On June 30, 1976, the inventory of unexpired clearances which includes statistical surveys or reports, program evaluation, management reports, and other recordkeeping requirements included about 3,300 repetitive reports and 500 single-time reports are carried out by contracts. About five percent of the repetitive surveys and about 50 percent of the single-time involved contractors. The problems of statistical data collection activities involving contractors outside of government are complex. In the development of better procedures for Phase II of the President's program, we have identified several things which should be done to improve contracting quality. They are:

1. The RFPs must clearly spell out the objectives of the survey and specify the statistical requirements to fulfill these objectives. The project officer writing the RFP needs both subject matter and statistical knowledge. If it is necessary to consult with a subject matter or a statistical expert, the project officer should do so. The definition of the programmatic goals of the data collection effort, however, are the responsibility of the sponsoring agency. Choosing the specific design alternatives must be worked out by the contractor.

2. Choosing among competitive RFPs requires sufficient insight to determine that the proposal submitted will fulfill the data needs and that the contractor has the required expertise.

3. Monitoring projects requires expertise in the sponsoring agency to assure that the data to be produced will address the issues at hand and meet the quality requirements needed.

4. An additional problem is that a principal contractor often subcontracts part of the work. Although this might be beneficial, if specific expertise is being sought, some negative consequences may ensue. The cost of the project may be increased by having profits at various levels of performance; in addition, further monitoring of the project is needed since each level of subcontractors should be monitored.

Future Areas of Investigation

In conclusion, two specific comments on future directives may be in order:

First, because the SPD is located within the Office of Management and Budget, we are very concerned with the cost of producing statistical data of adequate quality. During the clearance process, we often interact with appropriate budget examiners who will carefully evaluate costs of projects. Although we have not developed any specific guidelines of how much should be spent in collecting data using this or that methodology or implementing the project within the Federal sponsoring agency, by contract with another Federal agency or by contract with a noncommercial or commercial organization, a long-range project which would investigate factors which determine costs as well as how quality and costs of surveys are related would be very useful to us.

Second, the SPD is preparing "A Framework for Planning U.S. Federal Statistics, 1978-1989," in cooperation with the Federal statistical agencies. One chapter in the plan will address "Standards for Statistical Methodology." It will address some of these issues directly. For example:

-- A recommendation which may be included in the plan would suggest the revision of "Standards for Statistical Surveys" as given in Exhibit A of OMB Circular No. A-46. In addition, the development of standards for contracting statistical surveys would be recommended. This should include not only the standards for RFPs, but also standards for monitoring contractual data collection and processing.

-- Another recommendation might involve a centralized clearance system for publication of data. The objective of selective reviewing of new data releases in a centralized office before they are published would be to ensure a more uniform quality of Federal statistical data being released as well as to monitor the information given to users on the quality of the data. The review of the data after it is produced, in addition to the initial clearance of data collection projects, would give the SPD a much better capacity to control the quality of Federal statistical data.

Conclusion

Hence there is considerable evidence about the importance of this ASA project and related investigations. We are already taking some actions in OMB to improve the quality of government-funded surveys, and much activity is anticipated in the months ahead.

IDENTIFICATION AND ESTIMATION OF AGE, PERIOD AND COHORT EFFECTS IN THE ANALYSIS OF DISCRETE ARCHIVAL DATA

Stephen E. Fienberg, University of Minnesota
William M. Mason, University of Michigan

1. INTRODUCTION

The discussion below summarizes the extended version of the paper, which exceeds the page limit for inclusion in the Proceedings.

This paper shows how to identify and estimate age, period and cohort effects in models which are log-additive with respect to a categorical response variable, and discusses the source of the identification problem in such models. Mason et al. [1973] have a similar intent, but focus on models with quantitative response variables.

In the general case we treat there are replicated cross-sectional data for each of J regularly spaced points in time. Our formal considerations apply to a wide variety of data. For convenience of reference, however, we shall suppose that our data are from sample surveys of individuals. Then, for each of the J points in time (hereafter periods) the data can be combined into age groups of respondents, e.g., 20-24 years of age. We assume that the range in time (e.g., years) covered by each age group equals the interval in time between successive periods for which we have data. Thus, all those in the i-th of I age groups in the j-th of J periods correspond to the same birth cohort as those in age group i+1 at the subsequent period, j+1. With I age groups and J periods there are K = I+J-1 cohorts.

We are interested in the effects of age, period, and cohort on a categorical response variable. For simplicity we begin with a dichotomous response variable. Let P_{ijk} denote the probability of a positive response given age i, period j and cohort k = i-j+J. Then one possible model of interest is the logistic response model:

$$\Omega_{ijk} = \log \frac{P_{ijk}}{1 - P_{ijk}}$$

$$= W + W_{1(i)} + W_{2(j)} + W_{3(i-j+J)} \qquad (1)$$

where the subscripted parameters in (1) are deviations from W, i.e.,

$$\Sigma_i W_{1(i)} = \Sigma_j W_{2(j)} = \Sigma_k W_{3(k)} = 0. \qquad (2)$$

This model postulates simultaneous age, period and cohort effects on the log-odds (or logit) of the probability of success. The notation we use here is consistent with that in Bishop, Fienberg and Holland [1975]. The model can be expanded to include further explanatory variables (e.g., sex, race, socioeconomic status) as well as their interaction effects with age, period and cohort. The model is directly analogous to the age-period-cohort model for quantitative response variables discussed by Mason et al. [1973].

The problems of identification and estimation of model (1)-(2), and ways of arraying the data pertinent to model (1)-(2), are most simply illustrated for the case in which there are data from 3 periods, 3 age groups and, therefore, 5 cohorts. The extended version of the paper subjects the 3x3x2 case to detailed analysis, and then proceeds to the IxJx2 general case. Here, we shall illustrate one way of arraying the data for the case of 3 age groups and 3 periods, and then state the results for the general case in compressed form.

For the 3-age group and 3-period case, the basic data with which to formulate, estimate and assess the adequacy of an age-period-cohort model come from a series of 3 surveys, one for each period. The data then consist of counts $\{x_{ijk\ell}\}$, where $\ell = 1$ corresponds to a positive response and $\ell = 2$ to a negative response, and thus the counts form a 3x3x2 cross-classification with the sample sizes (marginal configurations $\{x_{+j++}\}$) fixed by design, as depicted in Table 1. It is important to bear in mind that one of the first three subscripts is redundant since k = i-j+3.

If we define the expected cell values corresponding to Table 1 under the logistic response model (1)-(2) by $\{m_{ijk\ell}\}$ we have:

$$m_{ijk1} = x_{ijk+}P_{ijk} \qquad (3)$$

and

$$m_{ijk2} = x_{ijk+}(1 - P_{ijk}) \ . \qquad (4)$$

The basic logistic response model can now be written in terms of expected cell values as:

$$\Omega_{ijk} = \log \frac{m_{ijk1}}{m_{ijk2}}$$

$$= W + W_{1(i)} + W_{2(j)} + W_{3(i-j+J)} \ , \qquad (5)$$

and analyses involving this model treat the marginal configuration $\{x_{ijk+}\}$ as fixed, even though only the totals $\{x_{+j++}\}$ are fixed by design.

Table 1 is just one of several ways in which the data can be arrayed. Under certain circumstances it may be preferable to construct age by cohort or period by cohort tables. Such alternative tables are also used in the extended version of the paper to aid the exposition of the identification problem.

2. IxJx2 CASE

A. Identification

Because cohort is determined uniquely by age and period, we must take care to ensure that the unique effects of age, period and cohort in the logistic response model are estimable. In the linear model for quantitative response variables analogous to (1)-(2) it is well known (Mason et al. [1973]) that all of the parameters are not estimable. The same is the case for the logistic response model.

In the case of I age groups and J periods it might be supposed that there are I-1 independent parameters for the effect of age on the response variable, J-1 for the period effects, and I+J-2 for the cohort effects for a total of 2I+2J-4. It turns out, however, that the number of independent effect parameters is 2I+2J-5, which is one less than specified by the basic logistic response model, (1)-(2). Thus, not all of the effect parameters specified by model (1)-(2) are identified.

The source of the identification problem can be described in various ways. One useful insight is that the effects of age, period and cohort contain linear and higher order components (e.g., quadratic components in the 3x3x2 case). It can be shown that the linear component of any one set of effect parameters (e.g., those of age) can not be separated from the linear components of the other two sets of parameters.

In order to estimate all of the independent effect parameters of model (1)-(2) it is necessary to put a single restriction on the model, e.g., $W_{1(1)} = W_{1(3)}$, $W_{2(1)}$ = constant, or $W_{3(1)} = W_{3(2)}$. We refer to such a restriction as an identification specification. Different identification specifications lead to effectively different models. An identification specification is like any other assumption in a statistical model that is not capable of direct verification as part of an analysis; it must be grounded in substantive theory relating to the data in question or it must come from observations on and analyses of other data on related phenomena.

Although the technical aim of making an identification specification is to allow the estimation of the linear components of the effect parameters, the most reasonable types of specifications are likely to be that two or more age groups, periods or cohorts have the same effects on the log-odds-ratios. What is more interesting from a substantive point of view than specifying a single identification specification, is the specification of overidentifying restrictions on the effect parameters based on considerable collateral information. When the resulting model fits the data well, not only do we solve the identification problem but we also get some verification of hypotheses related to the substantive theory.

B. Estimation

Given the logistic response model (1)-(2) and J independent simple random samples at 3 properly spaced points in time, and given an identification specification, it is possible to obtain maximum likelihood estimates of the effect parameters and of the expected values $\{m_{ijk\ell}\}$ corresponding to the observed frequencies $\{x_{ijk\ell}\}$. The likelihood equations can be solved by the Newton-Raphson iterative procedure (Bock [1975], Haberman [1974]). Alternatively, for sufficiently simple identification specifications the likelihood equations can be solved using iterative proportional fitting (Bishop, Fienberg and Holland [1975]), or for more complex identification specifications using generalized iterative proportional fitting (Darroch and Ratcliff [1972]). If (generalized) iterative proportional fitting is used, one solves first for the $\{\hat{m}_{ijk\ell}\}$ and then for the estimated effect parameters. If the Newton-Raphson method is used, one solves first for the estimated effect parameters. Thus, use of the Newton-Raphson method requires prior resolution of the identification problem. This method has been programmed for general purpose work with discrete data (Bock and Yates, [1973]) and is preferable to iterative scaling for several reasons. The extended version of the paper discusses estimation in more detail than is possible here.

C. Degrees of Freedom

Degrees of freedom equal the number of conditional log-odds for the response variable minus the number of independent effect parameters minus one (for the grand mean). For the "full" model, i.e., the model in which only one identification specification has been made, there are (I-2)(J-2) degrees of freedom. Table 2 lists the full model and the 7 possible reduced models (to be discussed below), and the associated minimal sufficient statistics and degrees of freedom. Goodman [1975] gives a similar table.

D. Goodness of Fit

Once we have estimated expected cell values, we can test the goodness-of-fit of model (1)-(2) using either the Pearson statistic,

$$X^2 = \Sigma\frac{(x_{ijk\ell} - \hat{m}_{ijk\ell})^2}{\hat{m}_{ijk\ell}} , \qquad (6)$$

or the likelihood ratio statistic,

$$G^2 = 2\Sigma x_{ijk\ell}\log\frac{x_{ijk\ell}}{\hat{m}_{ijk\ell}} . \qquad (7)$$

If the model is correct then either statistic is asymptotically distributed as a chi-square variate with degrees of freedom determined as described above.

E. Reduced Models

If the logistic response model with age, period and cohort effects, and with an associated identification specification, provides an acceptable fit to the data, then we would logically wish to explore whether only two sets of effects may suffice, i.e., whether we can equate one set of effects (age or period or cohort) to zero. Fitting such reduced models is a straightforward task with any computer program designed to fit standard loglinear models to multidimensional arrays (with or without structural zeros). There is no longer an identification problem when we deal with reduced models because there is no way for a linear component for one type of effect to become confounded with the other two types.

The fit of reduced models can be assessed using the standard goodness-of-fit statistics, (6) and (7), and we can compare the fit of the reduced models to the specified age-period-cohort models using the log-likelihood-ratio statistic for nested models, i.e., the conditional likelihood ratio test for the fit of the reduced model given that the age-period-cohort model is correct. In a similar fashion we can fit reduced models with only one set of effect parameters. Degrees of freedom and other information for reduced models are given in Table 2.

3. MULTIVARIATE QUALITATIVE RESPONSE MODELS

We have thus far dealt with logistic response models measuring the effects of age, period, and cohort on a dichotomous response variate. Here we consider two extensions of these models for polytomous response variates. Suppose the response variable has L categories, and the basic data array is the IxJxL age-period-response table (alternatively, the Ix(I+J-1)xL age-cohort-response table, or the Jx(I+J-1)xL period-cohort-response table). As before, we label the counts in the array using four subscripts, i.e., $\{x_{ijk\ell}\}$ where $\ell = 1,2,\ldots,$ L, and we denote the corresponding expected cell values by $\{m_{ijk\ell}\}$. When we had a dichotomous response variate we considered a model for the single logit structure, $\log(m_{ijk1}/m_{ijk2})$. Now that we have L categories for our response variable we would like to consider models for L-1 different logit structures. Two possible ways to define these are:

$$\text{(i)}\ \log\left(\frac{m_{ijk\ell}}{\sum_{\ell'>\ell} m_{ijk\ell'}}\right) \text{ for } \ell = 1,2,\ldots, L-1\ , \quad (8)$$

and

$$\text{(ii)}\ \log\left(\frac{m_{ijk\ell}}{m_{ijk\ell+1}}\right) \text{ for } \ell = 1,2,\ldots, L-1\ . \quad (9)$$

If we would like to fit models with the same parametric structure to the L-1 logits, and to have them correspond as a group to a loglinear model for the $\{m_{ijk\ell}\}$, then our choice would be (9). (Note that (9), in such circumstances, is equivalent to models with the same parametric structure for the L-1 logits

$$\text{(iii)}\ \log\left(\frac{m_{ijk\ell}}{m_{ijkL}}\right) \text{ for } \ell = 1,2,\ldots, L-1\ , \quad (10)$$

or any other set of L-1 logits involving the logarithm of the odds for pairs of expected values). The generalization of the logistic response model we would consider for (9) is:

$$\Omega_{ijk\ell} = \log\left(\frac{m_{ijk\ell}}{m_{ijk\ell+1}}\right)$$

$$= W^{(\ell)} + W^{(\ell)}_{1(i)} + W^{(\ell)}_{2(j)} + W^{(\ell)}_{3(i-j+J)}\ , \quad (11)$$

(for $\ell = 1,2,\ldots, L-1$),

where

$$\sum_i W^{(\ell)}_{1(i)} = \sum_j W^{(\ell)}_{2(j)} = \sum_k W^{(\ell)}_{3(k)} = 0\ . \quad (12)$$

All of the earlier results for the dichotomous response variate carry over immediately to this set of models as long as we make sure that (a) summations involving the fourth subscript run up to L instead of 2, and (b) the degrees of freedom listed in Table 2 are all multiplied by L-1.

If the L response categories are ordered and it makes substantive sense to think of the effects linking the response variable to age, period, and cohort as increasing linearly with the category number (e.g., the category number represents some latent variable), then we may wish to test for the equality of various effect parameters, e.g.,

$$W^{(\ell)}_{1(i)} = W^{*}_{1(i)} \quad \text{for } \ell = 1,2,\ldots, L-1\ . \quad (13)$$

Such reduced models can be handled without trouble using the methodology of loglinear models with ordered categories for some of the variables (see, e.g., Fienberg [1977]).

When the response categories have a natural order, e.g., educational attainment (grade school, high school, college, graduate school), the other choice of logits, in expression (8) may be preferable. The quantities $(m_{ijk\ell}/\sum_{\ell'>\ell} m_{ijk\ell'})$ are often referred to as continuation ratios, and they are of substantive interest in various fields. There is also a technical reason for working with the logits in expression (8). Let $P_{ijk\ell}$ be the probability of a response in category ℓ given age i and period j, where $\sum_\ell P_{ijk\ell} = 1$. Then, when the $\{x_{ijk\ell}\}$ consist of observations from IJ independent multinomial variables with sample sizes $\{x_{ijk+}\}$ and cell probabilities $\{P_{ijk\ell}\}$,

$$m_{ijk\ell} = x_{ijk+}\, P_{ijk\ell}\ , \quad (14)$$

so that

$$\frac{m_{ijk\ell}}{\sum_{\ell'>\ell} m_{ijk\ell'}} = \frac{P_{ijk\ell}}{\sum_{\ell'>\ell} P_{ijk\ell'}}\ . \quad (15)$$

We can write the multinominal likelihood functions as products of L-1 binomial likelihoods, the ℓ-th of which has sample sizes $\{\sum_{\ell' \geq \ell} x_{ijk\ell'}\}$ and cell probabilities $\{P_{ijk\ell}/\sum_{\ell' \geq \ell} P_{ijk\ell'}\}$. This means that if we use the method of maximum likelihood to estimate the parameters in the logistic response models

$$\log\left(\frac{m_{ijk\ell}}{\sum_{\ell' > \ell} m_{ijk\ell}}\right) = W^{(\ell)} + W^{(\ell)}_{1(i)} + W^{(\ell)}_{2(j)} + W^{(\ell)}_{3(i-j+J)} \quad , \qquad (16)$$

(for $\ell = 1,2,\ldots, L-1$) ,

subject to (12), then we can do the estimation separately for each logit model using methods applicable to dichotomous response variates, and we can simply add individual chi-square statistics to get an overall goodness-of-fit statistic for the set of models. Moreover, the observed binomial proportions

$$x_{ijk\ell}/\sum_{\ell' \geq \ell} x_{ijk\ell'} \ , \qquad \ell = 1,2,\ldots, L-1 \ , \qquad (17)$$

are asymptotically independent of each other so that we can assess the fit to the L-1 logit models, and various associated reduced models, independently.

For the logistic response models in (16) it might be of substantive interest to explore the equality of parameters across models as in expression (13). The estimated expected values for such models and the associated tests of fit can be handled with each by thinking in terms of a set of counts with 5 subscripts, $\{y_{ijk\ell t}\}$ where

$$y_{ijk\ell t} = \begin{cases} x_{ijk\ell} & \text{for } t = 1 \\ \sum_{\ell' > \ell} x_{ijk\ell'} & \text{for } t = 2 \ . \end{cases} \qquad (18)$$

Now, if we let $m^*_{ijk\ell t}$ be the expected value under model (16) corresponding to $y_{ijk\ell t}$, then we fit the L-1 models simultaneously by fitting a hierarchical loglinear model to the $\{y_{ijk\ell t}\}$ with minimal sufficient statistics $\{y_{ijk+t}\}$, $\{y_{i++\ell t}\}$, $\{y_{+j+\ell t}\}$, $\{y_{++k\ell t}\}$. If we restrict the model so that (13) holds we fit the loglinear model with minimal sufficient statistics $\{y_{ijk+t}\}$, $\{y_{i++\ell +}\}$, $\{y_{+j+\ell t}\}$, $\{y_{++k\ell t}\}$. We can also handle similar reduced models involving equality of period and cohort effects across the L-1 logistic response structures.

4. OTHER EXTENSIONS

In the detailed paper we consider the identification problem for a subset of the cases in which the age group intervals are not identical to the period intervals. We examine in particular the case of r-year (e.g., 5-year) age groups and 2r-year (e.g., 10-year) periods. Once an identification specification has been made for this case, estimation can proceed as described earlier.

5. EXAMPLE

The extended version of the paper includes a detailed analysis of the educational attainment of white males. The data are from the U.S. Decennial Censuses of 1940-1970. Education is treated as a set of continuation ratios, and an overidentified age-period-cohort model is fit to the various logged continuation ratios. Reduced models are also fit and discussed.

ACKNOWLEDGEMENT

We are grateful to Michael Battaglia for able and devoted computational assistance.

REFERENCES

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. [1975]. Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachusetts: The MIT Press.

Bock, R. D. [1975]. Multivariate Statistical Methods in Behavioral Research. New York: McGraw-Hill.

Bock, R. D., and Yates, G. [1973]. MULTIQUAL: Log-linear Analysis of Nominal or Ordinal Qualitative Data by the Method of Maximum Likelihood--A Fortran IV Program. Chicago: National Educational Resources.

Darroch, J. N., and Ratcliff, D. [1972]. "Generalized iterative scaling for log-linear models," The Annals of Mathematical Statistics, 43, 1470-80.

Fienberg, S. E. [1977]. The Analysis of Cross-Classified Data. Cambridge, Massachusetts: The MIT Press.

Goodman, L. A. [1975]. "A note on cohort analysis using multiplicative models and the modified multiple regression approach." Unpublished manuscript.

Haberman, S. J. [1974]. The Analysis of Frequency Data. Chicago: University of Chicago Press.

Mason, K. O., Mason, W. M., Winsborough, H. H. and Poole, W. K. [1973]. "Some methodological issues in cohort analysis of archival data," American Sociological Review, 38, 242-58.

Table 1: Age by Period Display

Positive Response

Age	Period 1	2	3
1	x_{1131}	x_{1221}	x_{1311}
2	x_{2141}	x_{2231}	x_{2321}
3	x_{3151}	x_{3241}	x_{3331}

Negative Response

Age	Period 1	2	3
1	x_{1132}	x_{1222}	x_{1312}
2	x_{2142}	x_{2232}	x_{2322}
3	x_{3152}	x_{3242}	x_{3332}

Table 2: Information Associated with Model (1), and Various Reduced Models

Subscripted Logistic Parameters in Model (1)	Degrees of Freedom	Minimal Sufficient Statistics*
1. None	IJ-1	$\{x_{+++\ell}\}$
2. Age	I(J-1)	$\{x_{i++\ell}\}$
3. Period	J(I-1)	$\{x_{+j+\ell}\}$
4. Cohort	(I-1)(J-1)	$\{x_{++k\ell}\}$
5. Age, Period	(I-1)(J-1)	$\{x_{i++\ell}\}$, $\{x_{+j+\ell}\}$
6. Age, Cohort	(I-1)(J-2)	$\{x_{i++\ell}\}$, $\{x_{++k\ell}\}$
7. Period, Cohort	(I-2)(J-1)	$\{x_{+j+\ell}\}$, $\{x_{++k\ell}\}$
8. Age, Period, Cohort	(I-2)(J-2)	$\{x_{i++\ell}\}$, $\{x_{+j+\ell}\}$, $\{x_{++k\ell}\}$

*For each model we always include the totals, $\{x_{ijk+}\}$, implied by the logistic structure, as well as the statistics listed.

DISCUSSION

Halliman H. Winsborough, University of Wisconsin, Madison

In a number of places, Fienberg and Mason observe that the identification problem in cohort analysis pertains to the linear component of the effects of age, period and cohort on the dependent variable. Although this fact is relatively obvious once one has seen it, arriving at an understanding of the point has caused me some pain. It is perhaps useful, therefore, to discuss an alternative approach to the observation.

The identification problem was, in recent times at least, posed for demographers and sociologists by Ryder who observed that if the cohort variable is expressed as year of birth, age in years, and period by a date, then period equals age plus cohort. Thus, it does not make any sense to run a regression on, say, fertility rates of the form:

$$F = a+b_1A+b_2P+b_3C+e$$

where all variables are thought to be continuous, the reason being, of course, that the X'X matrix is not of full rank. The whole business of using dummy variables and more recently log-linear analysis for such problems has been an attempt to deal with this difficulty and with the fact that one would not expect the effects of age, period, or cohort to be linear anyway.

Given the latter observation, it is useful to investigate what would happen if we permitted nonlinear effects in the continuous variables model. Suppose, for example, we thought that

$$Y = a+b_1A+b_2P+b_3C+B_4A^2+b_5P^2+b_6C^2+e.$$

Here it is still the case that P = A+C. Thus, we will not be able to separate the three linear terms. We could, however, estimate

$$Y = a+(b_1+b_2)A+(b_3+b_2)C+b_4A^2+b_5P^2+b_6C^2+e$$

by substituting the identity for P. We have no difficulty estimating b_4, b_5 and b_6 since the linear identity yields

$$P^2 = A^2+2AC+C^2$$

and substitution gives

$$Y=a+(b_1+b_2)A+(b_3+b_2)C+(b_4+b_5)A^2+(b_6+b_5)C^2+2b_5AC+e.$$

Thus, substitution of the squared identity yields an equation in only A and C from which b_4, b_5 and b_6 are retrievable. A similar retrieval is possible for higher-order polynomials in P, A and C.

Insofar, therefore, as one generates a cohort model of the form

$$Y = f(P)+g(A)+h(C)+e$$

and insofar as f, g and h can usefully be approximated by a Taylor series, the identification problem exists only with the linear terms in each series. Setting any one of the three coefficients to zero will allow estimation of all higher-order terms.

What does this fact imply for exploratory cohort analysis, i.e., analysis of a substantive problem for which it is not possible to make a strong a priori assertion about the process--assertions such as Fienberg and Mason are able to creditably make with their education example? It implies that the second and all higher derivatives of f, g and h can be uniquely estimated but the first derivative cannot. (This is the rationale behind the somewhat cryptic observation in the Fienberg - Mason paper that the "acceleration" of a variable with age, period, or cohort is estimable.) An analogue of this assertion holds for discrete coding of the independent variables. For the dummy variable method of analysis, Mason et al. [1] present an example showing that various choices of identifying restriction yield quite different effect parameters. Inspection of that example shows that the first differences of these parameters also vary with identifying restrictions. But second and all higher-order differences are invariant.

If one knows rather little about the process he is investigating, this estimability of second and higher-order differences in effect parameters is, I suppose, better than nothing. It certainly implies that such cohort models are not "hopelessly" under-identified, if that sometimes emotional phrase is meant to suggest that exploratory cohort analysis is inherently doomed to be completely unproductive.

There is a second point about the estimability problem in cohort analysis which seems to me important to make. The point is that the problem is not inevitable. It is not inevitable because Ryder's identity that period equals age plus cohort is not true for many data structures. Consider respondents born in the year 1900 and interviewed on July 1, 1976. If these respondents are asked their age, about half of them will correctly respond that they are 75, while the rest will correctly declare they are age 76. These people born on or before July 1, 1900 are 76; those born after July 1, 1900 are 75. This fact is, of course, familiar to demographers and is represented in that field by the Lexis diagram. Thus, if one's model posits a set of effects associated with number of whole-life-years lived and a set associated with period of interview, and if interviews do not occur on the first of January, the Ryder identity simply does not hold. Period will, of course, have quite a high multiple correlation with age and cohort but the resultant multi-colinearity is conceptually rather different than the estimability problem.

It can be argued that the above assertion represents a kind of "trick." If variables were scaled continuously, and thus accurately represented the "real" continuous nature of time, it can be argued, the identity would hold. There are two reasons why such an assertion of the greater reality of continuity is as "tricky" as my assertion. First, at the limit, it is difficult to believe that birth is an instantaneous process. It seems more reasonable to regard birth as a process occupying a time interval with any assignment of days, hours, minutes and seconds to the event as a crude way of locating the interval in time. Second, many of the dependent variables of interest represent some measure of a process which occurs in an interval. Examples are whether or not a child was born in the last year, earnings in the last year, or hours worked last week. Thus, it seems to me the argument that the Lexis diagram escape from the trap is a trick is based on an argument about the greater "reality" of continuity which is itself a trick. Of course, taking advantage of the Lexis diagram to ameliorate the problem requires that the analyst have great control over his data. Particularly it requires information which can yield both year of birth and current age. For many problems in the analysis of archival data such control is not available. In these situations, the estimability problem will exist.

REFERENCES

[1] Mason, Karen Oppenheim and William M. Mason, H. H. Winsborough, W. Kenneth Poole. "Some Methodological Issues in Cohort Analysis of Archival Data." American Sociological Review 38 (April): 242-258. 1973.

COEFFICIENTS OF MULTIPLE ASSOCIATION FOR RANKED DATA

Alan Agresti, University of Florida

1. INTRODUCTION

In this article, we propose a coefficient of multiple rank association $\tau_{Y\cdot\underline{X}} = \tau_{Y\cdot X^{(1)},\ldots,X^{(k)}}$ for describing a specific aspect of the association between a dependent variable Y and a set of independent variables $(X^{(1)}, X^{(2)}, \ldots, X^{(k)})$, when the available statistics consist of rankings on these variables for a sample of observations. In the next section $\tau_{Y\cdot\underline{X}}$ is defined as a generalization of Kendall's tau for two variables, in that it is based on the orderings for pairs of observations on each of the variables. The measure may be interpreted as a weighted average of the absolute values of the Kendall's taus between Y and each $X^{(i)}$ over sets of pairs with fixed orderings on $\{X^{(\nu)}, \nu = 1,2,\ldots,k\}$. In addition, $\tau_{Y\cdot\underline{X}}$ has a proportional reduction in error interpretation based on predicting pairwise ordering on Y using pairwise orderings on $\{X^{(\nu)}\}$. The extent of the increase in value of $\tau_{Y\cdot\underline{X}}$ as additional independent variables are added to the system is thus a measure of the improvement in predictive ability of these pairwise orderings on Y, for the given prediction rule. An additional coefficient $\tau^{(2)}_{Y\cdot\underline{X}}$, based on a different prediction rule, is also considered and seen to be more useful that $\tau_{Y\cdot\underline{X}}$ when $k = 2$.

In Section 3, the coefficients are generalized for application when there are some tied ranks, or the variables are ordinal caregorical in nature. The measure $\tau'_{Y\cdot\underline{X}}$ is defined in terms of all pairs of observations untied with respect to Y, and is seen to have similar properties as the coefficient $\tau_{Y\cdot\underline{X}}$ defined for the full-rank (no ties) case.

Examples of the calculation of these measures are given in a corresponding technical report (see Agresti (1976)). Also, asymptotic sampling distributions are considered in the report, as well as comparisons with other ordinal measures of multiple association which have been formulated.

2. MULTIPLE TAU COEFFICIENTS

The multiple tau coefficients which we shall define in this section are based on a generalization of the proportional reduction in error interpretation for the absolute value of Kendall's tau (denoted by τ_{YX} for the population value) between two variables Y and X (see, e.g., Ploch (1974)). For this bivariate case, let C and D represent the numbers of concordant and discordant pairs of observations, and suppose that there are no tied pairs.

If one were to predict at random for each of the $n(n-1)/2$ pairs of observations whether that pair was concordant or discordant (i.e., for each pair, predict concordance with probability 1/2, predict discordance with probability 1/2), the expected number of prediction errors would be $(C+D)/2 = n(n-1)/4$. If, on the other hand, one knows that $\tau_{YX} > 0$ and predicts concordance for each pair, the number of errors would be D. This results in a proportional reduction in error of

$$\frac{n(n-1)/4 - D}{n(n-1)/4} = \frac{C-D}{n(n-1)/2} = \tau_{YX} \qquad (2.1)$$

Similarly, if one knows $\tau_{YX} < 0$ and always predicts discordance, the proportional reduction in error is

$$\frac{D-C}{n(n-1)/2} = |\tau_{YX}|.$$

One could interpret $|\tau_{YX}|$, then, as the proportional reduction in error which results from predicting the ordering of pairs of observations on Y, based on having knowledge of the orderings of the pairs on X (and using the rule whereby the majority ordering is always predicted), relative to possessing no information about the orderings on X.

2.1 Definition of $t_{Y\cdot\underline{X}}$

Now, suppose that we wish to describe the association between a dependent variable Y and a collection of independent variables $\underline{X} = (X^{(1)},\ldots,X^{(k)})$. We shall next construct a similar type of coefficient with predictions of the ordering on Y based on the orderings on the $\{X^{(\nu)}, \nu = 1,\ldots,k\}$ for each pair of observations. In this section, for simplicity, we shall assume that there are no tied pairs with respect to any of the variables.

Let $(Y_1,\ldots,Y_n)$, $(X_1^{(1)},\ldots,X_n^{(1)}),\ldots,(X_1^{(k)},\ldots,X_n^{(k)})$ denote the rankings on $Y, X^{(1)},\ldots,X^{(k)}$ for n observations in some sample, and for a pair of observations (i,j), let

$$S_\nu(i,j) = S\left[(Y_j - Y_i)(X_j^{(\nu)} - X_i^{(\nu)})\right], \; \nu = 1,\ldots,k \qquad (2.2)$$

where S is the sign function

$$S[u] = \begin{cases} -1, & u < 0 \\ 0, & u = 0 \\ 1, & u > 0 \end{cases}$$

Also, denote $(S_1(i,j),\ldots,S_k(i,j))$ by $\underline{S}(i,j)$, and let

$$A(\underline{\delta}) = A(\delta_1,\ldots,\delta_k) = \{(i,j) : \underline{S}(i,j) = \underline{\delta}\}.$$

For example, $A(1,1,\ldots,1)$ is the set of pairs of observations which are simultaneously concordant between Y and each $X^{(\nu)}$, $\nu=1,\ldots,k$. If the pair of observations (i,j) is in $A(\underline{\delta})$, then that pair is $Y-X^{(\nu)}$ concordant (discordant) if $\delta_\nu = 1$ ($\delta_\nu = -1$). Notice that the $\{A(\delta_1,\ldots,\delta_k),\ \delta_\nu = \pm 1,\ \nu=1,\ldots,k\}$ create a partition of the $n(n-1)/2$ pairs of observations. Let

$$D_k = \{(\delta_1,\ldots,\delta_k):\ \delta_\nu = \pm 1,\ \nu=1,\ldots,k\}, \tag{2.3}$$

and denote by $N(\underline{\delta}) = N(\delta_1,\ldots,\delta_k)$ the number of pairs of observations in the set $A(\underline{\delta})$, so that

$$\sum_{D_k} N(\underline{\delta}) = n(n-1)/2$$

Now, for each element $\underline{\delta}$ of D_k, $A(\underline{\delta}) \cup A(-\underline{\delta})$ is the set of pairs of observations with a certain fixed ordering on the $\{X^{(\nu)},\ \nu=1,\ldots,k\}$, namely

$$\begin{array}{ll} X^{(u)} - X^{(w)} & \text{concordant if } \delta_u\delta_w = 1 \\ X^{(u)} - X^{(w)} & \text{discordant if } \delta_u\delta_w = -1, \\ & 1 \le u \le w \le k. \end{array}$$

There are several ways in which one could form predictions for the Y ordering for pairs in this set. The prediction rule for the coefficient to which we shall devote primary attention specifies that for each set $A(\underline{\delta}) \cup A(-\underline{\delta})$ of pairs with fixed orderings on the $\{X^{(\nu)}\}$, one should predict ordering on Y such that

$$\begin{array}{l} \underline{S}(i,j) = \underline{\delta} \text{ if } N(\underline{\delta}) \ge N(-\underline{\delta}) \\ \underline{S}(i,j) = -\underline{\delta} \text{ if } N(\underline{\delta}) < N(-\underline{\delta}) \end{array} \tag{2.4}$$

That is, if the majority of pairs in $A(\underline{\delta}) \cup A(-\underline{\delta})$ are $Y-X^{(\nu)}$ concordant (discordant), predict the orderings on Y for pairs in this set corresponding to $Y-X^{(\nu)}$ concordance (discordance).

According to this prediction rule, the number of prediction errors for pairs in $A(\underline{\delta}) \cup A(-\underline{\delta})$ is $\min[N(\underline{\delta}),N(-\underline{\delta})]$. On the other hand, random predictions of $Y-X^{(\nu)}$ concordance or discordance (with probability 1/2 for each) for pairs in $A(\underline{\delta}) \cup A(-\underline{\delta})$ would correspond to an expected number of errors of $[N(\underline{\delta}) + N(-\underline{\delta})]/2$. When predictions are considered over all pairs in all such sets with fixed $\{X^{(\nu)}\}$ ordering, the proportional reduction in error obtained from utilizing knowledge of ordering on the $\{X^{(\nu)}\}$ is

$$\begin{aligned} t_{Y\cdot\underline{X}} &= \frac{\sum_{D_k}[N(\underline{\delta}) + N(-\underline{\delta})]/2 - \sum_{D_k}\min[N(\underline{\delta}),N(-\underline{\delta})]}{\sum_{D_k}[N(\underline{\delta}) + N(-\underline{\delta})]/2} \\ &= \frac{n(n-1)/2 - \sum_{D_k}\min[N(\underline{\delta}),N(-\underline{\delta})]}{n(n-1)/2} \\ &= \frac{\frac{1}{2}\sum_{D_k}|N(\underline{\delta}) - N(-\underline{\delta})|}{n(n-1)/2} \end{aligned} \tag{2.5}$$

The factor of ½ occurs here and in some subsequent formulas due to the fact that both $|N(\underline{\delta}) - N(-\underline{\delta})|$ and $|N(-\underline{\delta}) - N(\underline{\delta})|$ occur in these sums when D_k is used as the index set.

Notice that $t_{Y\cdot\underline{X}}$ may be written as

$$t_{Y\cdot\underline{X}} = \sum_{D_k}\lambda(\underline{\delta})\ \frac{|N(\underline{\delta}) - N(-\underline{\delta})|}{N(\underline{\delta}) + N(-\underline{\delta})} \tag{2.6}$$

where $\lambda(\underline{\delta}) = \dfrac{N(\underline{\delta})}{n(n-1)/2}$ is the proportion of the $n(n-1)/2$ pairs of observations which are in $A(\underline{\delta})$. Letting

$$t(\underline{\delta}) = [N(\underline{\delta}) - N(-\underline{\delta})]/[N(\underline{\delta}) + N(-\underline{\delta})],$$

we see that

$$t_{Y\cdot\underline{X}} = \tfrac{1}{2}\sum_{D_k}(\lambda(\underline{\delta}) + \lambda(-\underline{\delta}))\,|t(\underline{\delta})| \tag{2.7}$$

is a weighted average of the absolute values of Kendall's tau type measures calculated within each set $A(\underline{\delta}) \cup A(-\underline{\delta})$ of orderings on the $\{X^{(\nu)}\}$. Since the joint orderings of the $\{X^{(\nu)}\}$ are fixed within $A(\underline{\delta}) \cup A(-\underline{\delta})$, $|t(\underline{\delta})|$ is in fact the absolute value of Kendall's tau between Y and each of the $X^{(\nu)}$ ($\nu=1,\ldots,k$), for that set of pairs.

The calculation of a coefficient such as $t_{Y\cdot\underline{X}}$ is based typically on a sample from some real or conceptual population of interest. Letting $P(\underline{\delta})$ denote the proportion of pairs of observations in $A(\underline{\delta})$ in this population, the corresponding population value of this coefficient is

$$\tau_{Y\cdot\underline{X}} = \tfrac{1}{2}\sum_{D_k}|P(\underline{\delta}) - P(-\underline{\delta})|. \tag{2.8}$$

Alternatively, let

$$D_M = \{\underline{\delta}:\ P(\underline{\delta}) > P(-\underline{\delta})\},\quad D_m = \{\underline{\delta}:\ P(\underline{\delta}) < P(-\underline{\delta})\}. \tag{2.9}$$

Then, for simplicity, we could rewrite the population coefficient as

$$\tau_{Y\cdot\underline{X}} = P_M - P_m = \sum_{D_M}[P(\underline{\delta}) - P(-\underline{\delta})], \tag{2.10}$$

where $P_M = \sum_{D_M} P(\underline{\delta}) = \Pr[S(i,j) \text{ in } D_M]$ and

$$P_m = \Pr[S(i,j) \text{ in } D_m]$$

for a randomly selected pair (i,j). We shall refer to the set of pairs indexed by D_M as those with majority ordering on Y with respect to $\{X^{(\nu)}\}$, and by D_m as those with minority ordering on Y with respect to $\{X^{(\nu)}\}$. Thus $\tau_{Y\cdot\underline{X}}$ is also similar in structure to Kendall's tau in that it may be interpreted as the difference in the probabilities of two types of pairs of observations.

2.2 Properties

We shall next consider some of the basic properties of $t_{Y\cdot\underline{X}}$. It is clear from the definition that $t_{Y\cdot\underline{X}}$ is invariant under order-preserving transformations on any of the variables. In the simple bivariate case,

$$t_{Y\cdot X} = \frac{|N(1) - N(-1)|}{n(n-1)/2} = \frac{|C - D|}{n(n-1)/2} \quad (2.11)$$

equals the absolute value of Kendall's tau between Y and X, the difference between the proportions of concordant and discordant pairs of observations. In the trivariate case,

$$t_{Y\cdot X^{(1)},X^{(2)}} = \{|N(1,1) + N(-1,-1)| + |N(1,-1) - N(-1,1)|\}/n(n-1)/2$$

$$= \frac{2}{n(n-1)} \max\{|(N(1,1) + N(1,-1)) - (N(-1,1) + N(-1,-1))|, |(N(1,1) + N(-1,1)) - (N(1,-1) + N(-1,-1))|\}$$

$$= \max\{|t_{YX^{(1)}}|, |t_{YX^{(2)}}|\} \quad (2.12)$$

The behavior of $t_{Y\cdot \underline{X}}$ becomes less trivial when the number of independent variables k exceeds two, as the simultaneous predictive power available from $(X^{(1)},\ldots,X^{(k)})$ may exceed that of the one most strongly associated with Y. In general, $t_{Y\cdot \underline{X}}$ is monotone increasing as the set of independent variables increases in size. To see that $t_{Y\cdot X^{(1)},\ldots,X^{(k)}} \le t_{Y\cdot X^{(1)},\ldots,X^{(k+1)}}$, we need only note that the partition of pairs into sets $\{A(\delta_1,\ldots,\delta_{k+1}) \cup A(-\delta_1,\ldots,-\delta_{k+1})\}$ with similar orderings on the $\{X^{(\nu)}, \nu = 1,\ldots,k+1\}$ is a subdivision of the partition $\{A(\delta_1,\ldots,\delta_k) \cup A(-\delta_1,\ldots,-\delta_k)\}$, and thus $\sum_{D_k} |N(\delta_1,\ldots,\delta_k) - N(-\delta_1,\ldots,-\delta_k)| = \sum_{D_k} |N(\delta_1,\ldots,\delta_k,1) + N(\delta_1,\ldots,\delta_k,-1) - N(-\delta_1,\ldots,-\delta_k,1) - N(-\delta_1,\ldots,-\delta_k,-1)|$

$$\le \sum_{D_k} |N(\delta_1,\ldots,\delta_k,1) - N(-\delta_1,\ldots,-\delta_k,-1)| + \sum_{D_k} |N(\delta_1,\ldots,\delta_k,-1) - N(-\delta_1,\ldots,-\delta_k,1)|$$

$$= \sum_{D_{k+1}} |N(\delta_1,\ldots,\delta_{k+1}) - N(-\delta_1,\ldots,-\delta_{k+1})|.$$

Notice that $t_{Y\cdot X^{(1)},\ldots,X^{(k)}} = t_{Y\cdot X^{(1)},\ldots,X^{(k+1)}}$ if and only if for each choice of $(\delta_1,\ldots,\delta_k)$, either

$$N(\delta_1,\ldots,\delta_k,1) \ge N(-\delta_1,\ldots,-\delta_k,-1)$$
$$\text{and } N(\delta_1,\ldots,\delta_k,-1) \ge N(-\delta_1,\ldots,-\delta_k,1),$$

or (2.13)

$$N(\delta_1,\ldots,\delta_k,1) \le N(-\delta_1,\ldots,-\delta_k,-1)$$
$$\text{and } N(\delta_1,\ldots,\delta_k,-1) \le N(-\delta_1,\ldots,-\delta_k,1);$$

that is, if the refinement in the partition of pairs by adding $X^{(k+1)}$ to the system does not result in a change in the predictions of $Y - X^{(\nu)}$ concordance or discordance for any of the pairs, $\nu = 1,\ldots,k$. In particular, if $|t_{X^{(\ell)}X^{(k+1)}}| = 1$ for some ℓ $(1 \le \ell \le k)$, then the partition is unchanged and $t_{Y\cdot X^{(1)},\ldots,X^{(k)}} = t_{Y\cdot X^{(1)},\ldots,X^{(k+1)}}$.

2.3 A Coefficient Based on a Different Prediction Rule

A rather striking property of $t_{Y\cdot \underline{X}}$ is that for $k = 2$, the reduction in prediction error equals just that corresponding to the more strongly associated $X^{(\nu)}$ of the two. Thus, $t_{Y\cdot \underline{X}}$ is a mathematically convenient but practically trivial measure when there are only two independent variables. The reason for this behavior lies with the prediction rule employed in formulating the coefficient. The rule of "predicting the majority ordering on Y with respect to the $\{X^{(\nu)}\}$" is a very simple one which leads to an easily interpretable coefficient. However, there is nothing unique about it, and more complex rules are necessary to produce a non-trivial measure when $k = 2$.

To formulate alternative coefficients of a nature similar to $t_{Y\cdot \underline{X}}$, one need only change the prediction rule. For example, suppose that a "proportional" prediction rule is utilized. That is, for a pair of observations in $A(\underline{\delta}) \cup A(-\underline{\delta})$, predict that

$$\underline{S}(i,j) = \underline{\delta} \text{ with probability } N(\underline{\delta})/(N(\underline{\delta}) + N(-\underline{\delta}))$$
$$\underline{S}(i,j) = -\underline{\delta} \text{ with probability } N(-\underline{\delta})/(N(\underline{\delta}) + N(-\underline{\delta})). \quad (2.14)$$

Then, the expected number of prediction errors for all such pairs in $A(\underline{\delta}) \cup A(-\underline{\delta})$ is $2N(\underline{\delta})N(-\underline{\delta})/(N(\underline{\delta}) + N(-\underline{\delta}))$, and considered over all such sets with fixed $\{X^{(\nu)}\}$ orderings, the proportional reduction in error using this rule is

$$t^{(2)}_{Y\cdot \underline{X}} = \frac{n(n-1)/2 - 2\sum_{D_k} N(\underline{\delta})N(-\underline{\delta})/(N(\underline{\delta}) + N(-\underline{\delta}))}{n(n-1)/2}. \quad (2.15)$$

It is easily verified that when $k = 1$, this coefficient reduces to the square of Kendall's tau (see Ploch (1974)). As the set of independent variables increases in size, $t^{(2)}_{Y\cdot \underline{X}}$ remains constant if the relative proportions used in the prediction remain unchanged. For example, $t^{(2)}_{Y\cdot X^{(1)},\ldots,X^{(k)}} = t^{(2)}_{Y\cdot X^{(1)},\ldots,X^{(k+1)}}$ if for all $\underline{\delta}$ in D_k

$$\frac{N(\delta_1,\ldots,\delta_k)}{N(-\delta_1,\ldots,-\delta_k)} = \frac{N(\delta_1,\ldots,\delta_k,1)}{N(-\delta_1,\ldots,-\delta_k,-1)} = \frac{N(\delta_1,\ldots,\delta_k,-1)}{N(-\delta_1,\ldots,-\delta_k,1)} \quad (2.17)$$

In particular, $t^{(2)}_{Y\cdot X^{(1)},X^{(2)}} > \max\left[t^{(2)}_{Y\cdot X^{(1)}}, t^{(2)}_{Y\cdot X^{(2)}}\right]$ unless

$$\frac{N(1)}{N(-1)} = \frac{N(1,1)}{N(-1,-1)} = \frac{N(1,-1)}{N(-1,1)}.$$

Thus, when $k = 2$, $t^{(2)}_{Y\cdot \underline{X}}$ is an especially useful measure of multiple rank association. For $k > 2$, $t^{(2)}_{Y\cdot \underline{X}}$ can be used as a supplementary measure to $t_{Y\cdot \underline{X}}$. However, it does not have quite as simple an interpretation as $t_{Y\cdot \underline{X}}$, and its value may seem somewhat artificial to the user, since it is naturally comparable to the squared rather that the unsquared Kendall's tau values. In essence, predictions based on this rule can result in no larger a reduction in error than those based on the rule previously described. Of course, $\sqrt{t^{(2)}_{Y\cdot \underline{X}}}$ could be used in comparison with the tau values, although then this coefficient lacks an interpretation. Also, the asymptotic moments and sampling distribution of $t^{(2)}_{Y\cdot \underline{X}}$ seem to be difficult to derive.

3. A MULTIPLE TAU COEFFICIENT FOR ORDINAL CATEGORICAL DATA

Tied pairs of observations would typically exist for most systems of variables in the social and behavioral sciences, where variables are commonly measured on ordinal categorical scales. If only a small proportion of pairs of observations are tied on at least one of the variables, one could continue to use $t_{Y\cdot \underline{X}}$ as defined in the previous section (tied pairs being omitted in the numerator). However, this results in a reduction in the potential magnitude of the measure, which becomes substantial as the proportion of tied pairs increases. For example, if the dependent variable is dichotomous with proportions .2 and .8 of observations in the two categories, the maximum possible value for $t_{Y\cdot \underline{X}}$ would be .32 (the proportion of pairs untied on Y) regardless of the distribution of ties among the independent variables.

3.1 Definition of $t'_{Y\cdot \underline{X}}$

To permit a maximum value of one and to ensure that the value does not decrease as independent variables are added to the system, one could base the coefficient on those pairs untied on Y and on at least one $X^{(\nu)}$, but standardize in the denominator according to the number of pairs untied on Y. That is, for $\delta_\nu = -1, 0$ or 1, $\nu = 1,\ldots,k$, let

$$A(\underline{\delta}) = \{(i,j):\ S[Y_j - Y_i] \neq 0 \text{ and } \underline{S}(i,j) = \underline{\delta}\} \quad (3.1)$$

and let $N(\underline{\delta})$ be the number of pairs of observations in $A(\underline{\delta})$. Let T_Y denote the number of pairs tied with respect to Y; i.e., if there are a_0 distinct values of Y with n_i observations tied at the i-th level, then $T_Y = \sum_{i=1}^{a_0} n_i(n_i - 1)/2$. Then, $\{A(\underline{\delta}),\ \delta_\nu = -1,0,1,\ \nu = 1,\ldots,k\}$ is a partition of the $n(n-1)/2 - T_Y$ pairs untied on Y, and $A(\underline{\delta}) \cup A(-\underline{\delta})$ is again the set of pairs with a fixed particular ordering on the $\{X^{(\nu)}\}$. Letting

$$D'_k = \{(\delta_1,\ldots,\delta_k):\ \delta_\nu = -1,\ 0, \text{ or } +1,\ \nu = 1,\ldots,k, \text{ at least one } \delta_\nu \neq 0\}, \quad (3.2)$$

we define

$$t'_{Y\cdot \underline{X}} = \frac{\frac{1}{2}\sum_{D'_k} |N(\underline{\delta}) - N(-\underline{\delta})|}{n(n-1)/2 - T_Y}. \quad (3.3)$$

Notice that $t'_{Y\cdot \underline{X}}$ may be rewritten as

$$t'_{Y\cdot \underline{X}} = \frac{\frac{1}{2}\left[\frac{n(n-1)}{2} - T_Y\right] - \frac{1}{2}\left[\sum_{D'_k} \min[N(\underline{\delta}), N(-\underline{\delta})] + N(0,\ldots,0)\right]}{\frac{1}{2}\left[\frac{n(n-1)}{2} - T_Y\right]} \quad (3.4)$$

That is, $t'_{Y\cdot \underline{X}}$ is the proportional reduction in error of predictions of the ordering on Y (for those pairs untied on Y) obtained by predicting majority ordering based on ordering of the $\{X^{(\nu)}\}$, relative to predicting randomly. Of course, when all $\delta_\nu = 0$, predictions in effect are also made randomly since the $\{X^{(\nu)}\}$ provide no predictive information, resulting in an expected number of errors of $N(0,\ldots,0)/2$.

Alternatively, $t'_{Y\cdot \underline{X}}$ may be written as

$$t'_{Y\cdot \underline{X}} = \sum_{D'_k} \lambda(\underline{\delta}) \frac{|N(\underline{\delta}) - N(-\underline{\delta})|}{N(\underline{\delta}) + N(-\underline{\delta})}$$
$$= \frac{1}{2}\sum_{D'_k} (\lambda(\underline{\delta}) + \lambda(-\underline{\delta}))\,|t(\underline{\delta})|, \quad (3.5)$$

where $\lambda(\underline{\delta}) = \frac{N(\underline{\delta})}{n(n-1)/2 - T_Y}$ is the proportion of the pairs of observations untied on Y which are in $A(\underline{\delta})$. Thus, $t'_{Y\cdot \underline{X}}$ may be interpreted as a weighted average of the absolute values of the Kendall's taus within each set $A(\underline{\delta}) \cup A(-\underline{\delta})$ of orderings on the $\{X^{(\nu)}\}$, where a weight of $\frac{N(0,\ldots,0)}{n(n-1)/2 - T_Y}$ (the proportion of those pairs untied on Y which are tied on all $X^{(\nu)}$) is given to a value of $0 = N(0,\ldots,0) - N(0,\ldots,0)$. Here, $|t(\underline{\delta})|$ is the absolute value of Kendall's tau between Y and each of the $X^{(\nu)}$ such that $\delta_\nu \neq 0$, within the set of pairs $A(\underline{\delta}) \cup A(-\underline{\delta})$.

If $P(\underline{\delta})$ denotes the proportion of pairs of observations in $A(\underline{\delta})$ and p_i denotes the proportion of pairs tied on Y at the i-th of a_0 sets of ties on Y in some population of interest, then the population value of the coefficient $t'_{Y\cdot \underline{X}}$ is

$$\tau'_{Y\cdot \underline{X}} = \frac{\frac{1}{2}\sum_{D'_k} |P(\underline{\delta}) - P(-\underline{\delta})|}{1 - \sum_{i=1}^{a_0} p_i^2}$$
$$= \frac{\sum_{D'_M} [P(\underline{\delta}) - P(-\underline{\delta})]}{1 - \sum_{i=1}^{a_0} p_i^2} \quad (3.6)$$

where $D'_M = \{\underline{\delta} \text{ in } D'_k: P(\underline{\delta}) > P(-\underline{\delta})\}$.

Alternative coefficients could again be formulated based on different prediction rules. For example, an extension of the proportional prediction rule discussed in the last section yields the multiple measure for ordinal categorical data,

$$t^{(2)'}_{Y\cdot\underline{X}} = \qquad (3.7)$$

$$\frac{[n(n-1)/2 - T_Y] - 2 \sum_{D'_k \cup (0,\dots,0)} N(\underline{\delta})N(-\underline{\delta})/(N(\underline{\delta}) + N(-\underline{\delta}))}{n(n-1)/2 - T_Y}.$$

3.2 Properties

Clearly, $t'_{Y\cdot\underline{X}}$ is invariant under strictly order preserving transformations on any of the variables. When there are no tied pairs with respect to any of the variables, $t'_{Y\cdot\underline{X}}$ reduces to the coefficient $t_{Y\cdot\underline{X}}$ discussed in Section 2. In the bivariate case, $t'_{Y\cdot X}$ reduces to Somers' d_{XY} (see Somers (1962)), a well-known asymmetric ordinal measure of association. When $k=2$, $t'_{Y\cdot X^{(1)},X^{(2)}}$ is likely to be not much larger than $\max\{|t'_{Y\cdot X^{(1)}}|, |t'_{Y\cdot X^{(2)}}|\}$, but there is not necessarily equality here due to the additional contribution in the numerator of pairs tied on $X^{(1)}$ but not on $X^{(2)}$ and Y, or of pairs tied on $X^{(2)}$ but not on $X^{(1)}$ and Y. Again, though $t'_{Y\cdot\underline{X}}$ if of primary interest when $k \geq 3$, and a coefficient such as $t^{(2)'}_{Y\cdot\underline{X}}$ is likely to be of greater practical use when $k=2$.

With the addition of $X^{(k+1)}$ to the set of independent variables,

$$t'_{Y\cdot X^{(1)},\dots,X^{(k)}} \leq t'_{Y\cdot X^{(1)},\dots,X^{(k+1)}},$$

since the denominator remains constant and the numerator can not decrease when the partition of pairs $\{A(\underline{\delta}) \cup A(-\underline{\delta})\}$ is refined.

For additional properties and a numerical example, see the related technical report on these measures (Agresti (1976)).

BIBLIOGRAPHY

[1] Agresti, A. (1976). Coefficients of Multiple Association for Ranked Data. Technical Report No. 112, University of Florida Department of Statistics.

[2] Ploch, D. R. (1974). Ordinal Measures of Association and the General Linear Model, in Hubert M. Blalock, ed., Measurement in the Social Sciences, Chicago: Aldine-Atherton, Ch. 12.

[3] Somers, R. H. (1962). A New Asymmetric Measure of Association for Ordinal Variables. American Sociological Review, 27, 799-811.

SELECTING THE MOST RELIABLE FROM A SET OF JUDGES: TOWARD A GENERAL SOLUTION

Sandra L. Aivano, Yale University
Domenic V. Cicchetti and Jacob Levine, V.A. Hospital, West Haven, Conn.

Introduction

Social scientists frequently study variables that can be measured only by means of ordinal rating scales. Since the quality of the data collected using such scales could greatly influence the results obtained, pilot tests are often run to insure that the ratings in the main investigation will be accurate and reliable. Within this context, researchers often face the task of selecting the most reliable judges or raters to participate in the study.

One pilot testing procedure frequently used to assess raters' reliability and to choose the most reliable among them is to compare the raters' judgments to those of an expert or experts (Lehmann, Ban, and Donald, 1965; Fleiss, Spitzer, and Burdock, 1965). Judges whose scores deviate greatly from the expert judgments are either eliminated or trained in the use of the rating scale. This technique is not always feasible, since it assumes that expert judges can be found for the variable in question and that they can agree among themselves. However, it is difficult to establish experts on the basis of training or experience for many variables (such as the one to be discussed in this paper). Furthermore, it is often impossible or impractical for experts to participate in reliability studies (Fleiss, Spitzer, and Burdock, 1965; Smith, 1974). In view of the problems with this procedure, it is clear that a method for choosing reliable judges is needed that does not depend on expert judgments.

Several techniques have been proposed for comparing the reliability of judges and identifying the most reliable among them. One method involves computing the intraclass correlation coefficient (R_I) for all subsets of R raters from the pool of R´ raters, and selecting the subset with the highest R_I value (Burdock, Fleiss, and Hardesty, 1963). Another technique is to compute the Spearman Rho between each rater's ranking of the subjects and a composite ranking reflecting the rankings by all the remaining raters (Smith, 1974).

A new technique for selecting the most reliable raters from a larger rater pool is presented in this paper. This method was developed in the context of a pilot study designed to establish the difficulty of a series of cartoons to be used in a later investigation. Preliminary analyses revealed that agreement on this variable was not, in general, above chance expectancy. Thus, we were faced with the problem of determining which raters from the initial group could rate the cartoons reliably. This technique bears similarities to a technique briefly mentioned by Smith (1974) of computing the intercorrelations between all possible rater pairs, converting these values using Fisher's $\underline{Z}$ function, and using the average Z value for each rater as an index of that rater's reliability relative to the other prospective raters. Rather than using intercorrelations, however, the method presented here employs weighted kappa due to Cohen (1968) with a standard error due to Fleiss, Cohen, and Everitt (1969). Weighted kappa is specifically designed to measure agreement when the data are ordinal. It is more appropriate than other measures of association for ordinal scales since it takes into consideration the amount of agreement expected by chance alone. (For a further discussion of the kappa statistics relative to other available statistics for assessing rater agreement with qualitative data, see Fleiss (1975).)

Method

Ten judges rated the difficulty of 30 cartoons selected from magazines. The raters were staff psychologists at the V.A. Hospital, West Haven, Connecticut and Yale University. The variable, Difficulty Level, was measured on a 5-point ordinal scale with the following categories: (1) "very easy"; (2) "easy"; (3) "average"; (4) "difficult"; and (5) "very difficult."

Results and Discussion

Agreement Statistics

The agreement for each of the R(R-1)/2 rater pairs was assessed using weighted kappa (Cohen, 1968; Fleiss, Cohen, and Everitt, 1969) with a continuous-ordinal weighting system (Cicchetti, 1972, 1976; Cicchetti and Allison, 1973) as defined below. Weights (W) were computed by the formula:

$$W = \frac{k-1}{k-1}, \frac{k-2}{k-1}, \ . \ . \ . \ \frac{k-(k-1)}{k-1}, \frac{k-k}{k-1} \qquad [1]$$

where k refers to the number of points on the scale. Weights range from $\frac{k-1}{k-1}$ or 1, for ratings in perfect agreement, to $\frac{k-k}{k-1}$ or 0, for those that are the maximum possible number of scale points apart. Weights for the various levels of partial agreement, $\frac{k-2}{k-1}, \ . \ . \ . \ \frac{k-(k-1)}{k-1}$, assume values between 0 and 1. Using this weighting system, the proportion of observed agreement (PO); proportion of expected or chance agreement (PC); the level of chance-corrected agreement, or kappa, i.e., (PO-PC)÷(1-PC); the Z value of kappa; and its level of statistical significance, were computed for each rater pairing.

A recent Monte Carlo study has revealed that a minimum sample size of approximately $2k^2$ is needed to obtain valid results with the kappa statistics (Cicchetti and Fleiss, 1976). Since the thirty cartoons do not constitute a sufficient sample for the 5-point scale of Difficulty Level, categories 1 and 2 and categories 4 and 5 were collapsed, and the data were reanalyzed on a 3-point scale. The results obtained with the 3-point scale were very similar to those obtained using the 5-point scale. Ten and eleven statistically significant kappa values ($p \leq .05$) were obtained using the 3-point and 5-point rating systems, respectively. Nine of these significant kappas were for the same rater pairs on both the 3-point and 5-point scales. For the sake of brevity, the kappa statistics for only the 3-point scale are reported in Table 1.

Ranking Systems

As indicated from the kappa statistics in Table 1, the agreement for most judge pairs (35 from a total of 45) was not above chance expectancy at $p \leq .05$. Despite this overall lack of agreement, we wished to identify the raters that were the most reliable. To accomplish this, we developed two systems for ranking the raters based on the significance of the kappa values obtained in the pairwise comparisons.

Ranking System 1 consists of assigning consecutive integer ranks to the R(R-1)/2 rater pairings according to the magnitude of their Z of kappa values, with a rank of 1 for the rater pair with the highest Z value, and a rank of R(R-1)/2 for the rater pairing with the lowest Z value. The ranks for the R-1 comparisons associated with each rater are summed to obtain a composite rank reflecting that rater's reliability relative to the remaining raters. Then the composite rank scores of the raters can be compared, and the raters with the lowest scores can be identified as the most reliable. The results obtained by applying Ranking System 1 to the kappa statistics in Table 1 are presented in Table 2.

While System 1 is an index of the *relative* standing of each rater, it does not take into account the absolute level of the Z values. However, the magnitude of the Z values is often of great importance to the researcher, since it is of little value to identify the most reliable judges if, say, none of the kappa values obtained is statistically significant or all of the kappas are highly significant. Thus, to provide additional information for selecting the most reliable raters, we developed a second ranking system.

Ranking System 2 utilizes the number of significant ($p \leq .05$) and approaching significant ($p \leq .10$) Z of kappa values among the R-1 kappas associated with each rater. The raters are first ranked according to the number of significant Z values among the R-1 comparisons associated with each of them. Raters with the same number of significant Z values are further differentiated by the number of Z values approaching significance. The results obtained by applying this ranking system to the kappa statistics for the 45 rater pairings from Table 1 are presented in Table 3.

Selecting the Most Reliable Raters

As can be seen from Tables 2 and 3, the rank orderings of raters produced by the two ranking systems are very similar. The rank correlation coefficient (Spearman's Rho) between the two orderings is 0.83 with $p = .002$. The researcher is at liberty to decide how to divide the raters into subsets of the most reliable and least reliable, using the information obtained from the two ranking systems. For the 10 raters of Difficulty Level, we felt that Raters 2, 5, 6, 7, 8, and 9 could be considered reliable, while Raters 1, 3, 4, and 10 were markedly less reliable. This division of the raters seems reasonable, since it splits the raters at one of the points of greatest difference in composite scores (in Table 2, between 9 and 3 who are 34 points apart), and eliminates those raters with only one significant Z value (in Table 3). The *same* six raters (2, 5, 6, 7, 8, and 9) are identified as the most reliable by both ranking systems.

Both ranking systems proposed above evaluate the raters on the basis of their agreement *with all other raters*. Another important consideration in selecting the most reliable subset of judges is the extent to which the judges selected agree *among themselves*. Table 4 presents the proportion of observed agreement (PO); proportion of chance or expected agreement (PC); and the significance level (p) of chance-corrected agreement (or Kappa), for rater pairs in the most reliable subset (Raters 2, 5, 6, 7, 8, and 9) and least reliable subset (Raters 1, 3, 4, and 10), respectively. A comparison of the two portions (A and B) of the table shows that the overall levels of agreement between the more reliable raters are much higher than those between the less reliable raters. The mean of the PO values for the most reliable raters is 0.70 compared to a mean of 0.59 for the less reliable raters. Of the 10 significant ($p \leq .05$) kappas among all 45 rater pairs, 7 are between raters in the most reliable subset, while only 1 significant kappa is found in the least reliable subset. Thus, the raters who are most reliable with respect to all raters in the pool are also highly reliable with respect to each other.

The method employed here to find the most reliable subset of the ten raters of cartoon Difficulty is suggested as a general technique for selecting the most reliable judges from a larger judge pool. Comparing each rater relative to all the others using ranking systems based on the levels of chance-corrected agreement, as measured by Z of kappa values, seems a reasonable approach to the problem. This technique is based on the assumption that raters with the highest agreement relative to all the raters are, indeed, the best suited to participate in later investigations.

Another approach to selecting raters is to choose those with the greatest agreement among themselves, disregarding their levels of agreement with the remaining raters in the pool. This might be accomplished by computing the kappa statistics for all possible rater pairs and averaging the Z of kappa values for each rater pairing among the raters in each possible R rater subset of the R′ rater pool. This technique is often impractical, however, since the researcher must either decide beforehand the size of the subset he wishes to select or he must compute average Z values for all the possible subsets of all sizes. Further, this technique could result in the selection of raters who might not be well suited for future studies. For example, raters who tended to give the highest possible score would agree highly among themselves, yet would probably not be using the scale properly.

A computer program has been written to implement the rater selection method discussed in this paper. For each possible rater pair the program computes *weighted* kappa statistics, if the data are *ordinal*, or *unweighted* kappa statistics (Cohen, 1960; Fleiss, Cohen, and Everitt, 1969), if the data are *nominal*. Further, if the sample size is sufficiently large, bias is assessed for each rater pair by means of the chi-squared statistic developed by McNemar (1947). Finally, rankings of the raters by the two ranking systems, as well as a number of summary tables, are also provided.

Summary

Researchers often face the task of identifying the most reliable subset of a given set of judges. Specifically, this problem arose when ten judges rated the difficulty of a series of cartoons on an ordinal scale. Kappa statistics and various ranking techniques were used to obtain the information necessary to select the most reliable judges from the original pool. The observed agreement, chance agreement, and the statistical significance of their difference were computed for each pair of judges. These data were summarized in several tables, using ranking systems for both the judge pairs and the individual judges. An important contribution of one interjudge ranking system is that it assigned to each judge a composite score reflecting the reliability of his ratings relative to that of all the others. Finally, a computer program was written for use in resolving this type of research problem.

Acknowledgments

This research was supported by the West Haven V.A. Hospital (MRIS 1416). The authors express their appreciation to Lorraine R. Gambino for her excellent typing of the manuscript and editorial assistance.

References

Burdock, E.I., Fleiss, J.L., and Hardesty, A.S. A new view of interobserver agreement. *Personnel Psychology*, 1963, *16*, 373-384.

Cicchetti, D.V. A measure of agreement between rank ordered variables. *Proceedings of the American Psychological Association*, 1972, *7*, 17-18.

Cicchetti, D.V. Assessing inter-rater reliability for rating scales: Resolving some basic issues. *British Journal of Psychiatry*, 1976, in press.

Cicchetti, D.V. and Allison, T. Assessing the reliability of scoring EEG sleep records: An improved method. *Proceedings and Journal of the Electrophysiological Technologists' Association*, 1973, *20*, 92-102.

Cicchetti, D.V. and Fleiss, J.L. Comparison of the null distributions of Weighted Kappa and the C Ordinal Statistic. *Applied Psychological Measurement*, 1976, in press.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, *20*, 37-46.

Cohen, J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 1968, *70*, 213-220.

Fleiss, J.L. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, 1975, *31*, 651-659.

Fleiss, J.L., Cohen, J., and Everitt, B.S. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, 1969, *72*, 323-327.

Fleiss, J.L., Spitzer, R.L., and Burdock, E.I. Estimating accuracy of judgment using recorded interviews. *Archives of General Psychiatry*, 1965, *12*, 562-567.

Lehmann, H.E., Ban, T.A., and Donald, M. Rating the rater. *Archives of General Psychiatry*, 1965, *13*, 67-75.

McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 1947, *12*, 153-157.

Smith, J.M. A new rater selection technique for use with behavioral rating scales. *Journal of Clinical Psychology*, 1974, *30*, 40-43.

TABLE 1

AGREEMENT BETWEEN RATER PAIRS FOR 10 RATERS ON DIFFICULTY LEVEL FOR 30 CARTOONS ON A 3-POINT SCALE

Rank	Rater Pair[1]	PO	PC	Kappa	Z of Kappa	p of Kappa
1	3, 10	.77	.59	.43	2.94	.003
2	6, 8	.72	.61	.28	2.83	.005
3	2, 8	.78	.64	.40	2.80	.005
4	4, 7	.75	.60	.38	2.73	.006
5	5, 6	.75	.64	.31	2.68	.007
6	2, 9	.77	.63	.38	2.57	.010
7	2, 5	.72	.60	.30	2.12	.034
8	7, 9	.63	.52	.23	2.08	.037
9	7, 8	.65	.55	.22	2.05	.040
10	1, 9	.77	.67	.30	1.98	.048
11	8, 9	.75	.66	.27	1.86	.062
12	3, 8	.72	.62	.25	1.79	.073
13	5, 8	.70	.61	.22	1.65	.098
14	1, 6	.63	.57	.14	1.65	.099
15	2, 6	.67	.60	.17	1.63	.104
16	1, 7	.60	.52	.16	1.61	.108
17	3, 5	.68	.59	.23	1.60	.109
18	5, 9	.68	.60	.21	1.60	.109
19	2, 7	.63	.55	.19	1.59	.112
20	6, 9	.63	.57	.14	1.52	.128
21	2, 4	.58	.51	.15	1.38	.166
22	4, 8	.57	.50	.13	1.34	.179
23	5, 7	.65	.58	.17	1.33	.183
24	2, 3	.67	.60	.17	1.14	.253
25	6, 7	.70	.65	.13	1.09	.276
26	5, 10	.65	.59	.15	1.03	.302
27	1, 8	.72	.67	.14	.94	.346
28	1, 5	.65	.60	.12	.91	.361
29	1, 3	.67	.62	.12	.88	.376
30	3, 7	.60	.55	.11	.88	.376
31	4, 5	.60	.55	.10	.88	.378
32	3, 6	.63	.60	.08	.77	.438
33	6, 10	.63	.60	.08	.77	.438
34	4, 9	.52	.48	.07	.74	.457
35	8, 10	.65	.62	.08	.56	.578
36	4, 10	.55	.52	.06	.55	.581
37	3, 9	.63	.61	.06	.42	.676
38	7, 10	.57	.55	.04	.29	.768
39	4, 6	.65	.64	.02	.13	.895
40	2, 10	.60	.60	.00	.00	1.00
41	3, 4	.52	.52	-.01	-.06	.951
42	1, 2	.63	.64	-.02	-.11	.914
43	1, 10	.60	.62	-.05	-.38	.704
44	1, 4	.45	.47	-.04	-.46	.644
45	9, 10	.57	.61	-.11	-.78	.437

[1]Rater pairs are ordered by Z of kappa values.

TABLE 2

RANKING SYSTEM 1

RANKING THE RATERS BY COMPOSITE RANK SCORE

Rater	Composite Score	Ranks for Rater Pairs Comprising Composite Score
8	134	2+3+9+11+12+13+22+27+35
5	168	5+7+13+17+18+23+26+28+31
7	172	4+8+9+16+19+23+25+30+38
2	177	3+6+7+15+19+21+24+40+42
6	185	2+5+14+15+20+25+32+33+39
9	189	6+8+10+11+18+20+34+37+45
3	223	1+12+17+24+29+30+32+37+41
1	253	10+14+16+27+28+29+42+43+44
4	272	4+21+22+31+34+36+39+41+44
10	297	1+26+33+35+36+38+40+43+45

TABLE 3

RANKING SYSTEM 2

RANKING THE RATERS BY THE NUMBER OF SIGNIFICANT ($p \leq .05$) AND APPROACHING SIGNIFICANT ($p \leq .10$) Z OF KAPPA VALUES

Rater	$p \leq .05$	$.05 < p \leq .10$	$p \geq .10$
8	3	3	3
9	3	1	5
2	3	0	6
7	3	0	6
5	2	1	6
6	2	1	6
1	1	1	7
3	1	1	7
4	1	0	8
10	1	0	8

TABLE 4

WEIGHTED KAPPA STATISTICS FOR RATER PAIRINGS OF THE MOST RELIABLE AND LEAST RELIABLE RATERS: PO, PC, AND THE LEVEL OF STATISTICAL SIGNIFICANCE OF KAPPA (p)

A. Pairings of the Most Reliable Raters

Rater		2	5	6	7	8
5	PO	.72				
	PC	.60				
	p	.034*				
6	PO	.67	.75			
	PC	.60	.64			
	p	.104	.007**			
7	PO	.63	.65	.70		
	PC	.55	.58	.65		
	p	.112	.183	.276		
8	PO	.78	.70	.72	.65	
	PC	.64	.61	.61	.55	
	p	.005**	.098+	.005**	.040*	
9	PO	.77	.68	.63	.63	.75
	PC	.63	.60	.57	.52	.66
	p	.010*	.109	.128	.037*	.062+

B. Pairings of the Least Reliable Raters

Rater		1	3	4
3	PO	.67		
	PC	.62		
	p	.376		
4	PO	.45	.52	
	PC	.47	.52	
	p	.644	.951	
10	PO	.60	.77	.55
	PC	.62	.59	.52
	p	.704	.003**	.581

+ = Significant at .10 level
* = Significant at .05 level
** = Significant at .01 level

LONG-RANGE ESTIMATES WITH MICROSIMULATION OF THE COSTS AND CASELOADS OF INCOME ASSISTANCE PROGRAMS

Jodie T. Allen and Raymond J. Uhalde, Mathematica Policy Research

The rapid growth in expenditures on transfer programs witnessed during the last 20 years has raised serious questions concerning the inevitability that income assistance spending will consume an increasing proportion of this nation's productive output. The Washington office of Mathematica Policy Research contracted to provide assistance to the Congressional Budget Office (CBO) in developing estimates of the costs and caseloads of the major income assistance programs in the year 2000. As stipulated by CBO, the procedure employed takes explicit account of the design of existing transfer programs; and the interactions among them; makes no ancillary assumptions, other than those noted below, with regard to changes in program design or the introduction of new programs; and is consistent with mid-range census projections of population size, demographic composition and household structure.

The findings of the study are summarized in a report prepared by the Congressional Budget Office for the Senate Committee on the Budget.[1] That report also provides a comparison of these findings with those produced by a trend-line extrapolation methodology.

This report presents a summary of the major findings and a brief description of the methodology employed in (1) demographically ageing the 1973 Current Population Survey to conform to the mid-range Census projections of population size and distribution and trends in household formation rates; (2) projecting non-transfer income according to stipulated assumptions with regard to real income growth and price inflation; (3) imputing certain transfer income sources; and (4) simulating program costs and caseloads for income-tested transfer programs including Supplemental Security Income, Aid to Families with Dependent Children, Food Stamps, and Medicaid.[2]

CREATING A YEAR 2000 DATA BASE

A necessary step in the production of long-range estimates of the costs and caseloads of transfer programs is the development of a data base that is representative of the demographic and economic characteristics of the population in the future year of interest. A comparative static ageing procedure is used to project the March 1973 Current Population Survey to the year 2000. The ageing procedure is described as comparative static because population data are provided at two points in time without following the movement from one point to another. Data for the first point in time (the base year) are obtained from the data source being used. Data for the year in which we are interested (the simulation year) are derived by statically ageing the base year data. Thus, particular families or persons are not followed through time from the base year to the simulation year, as they would be in dynamic simulation. Rather, a family observed in the base year is aged to represent a family of similar characteristics in the simulation year. What does change is the number of families or persons with a particular set of characteristics and the income they receive. The goal is to produce an aged sample which is as similar as possible to the sample that would be drawn by a survey taken in the simulation year.

Operationally, the ageing is accomplished in two computational steps using a set of multipliers for each step. The first step, demographic ageing, consists of adjusting the sample weights attached to each family and person so that a given sample observation represents a new number of persons or families consistent with the projected population in that class. This is done by applying a set of demographic multipliers to the sample weights. The control totals, from which the demographic multipliers are derived, are in turn developed from Census Bureau population and household projections. The demographic multipliers are designed to account for projected changes in the age, sex, and household type composition of the population. To the extent these variables are correlated with other demographic and behavioral characteristics of the population, the reweighting procedure will alter the related population characteristics in a manner which, in some areas, may not accord with independent predictions. Of particular importance to policy analysis of tax and transfer programs is the labor supply behavior and earnings capabilities of the population. Given the historical and projected downward trend in birthrates, sole reliance on a reweighting procedure to achieve the average family size implied by the census projections would mean a relatively large inflation of the weights of small, base-year sample families. These families exhibit, among other characteristics, relatively high female labor force participation and earnings levels. Whether the increase in female labor force participation, generated indirectly by reweighting according to the number of children, would reconcile with independent projections is unclear. Consequently, the population of children was created through stochastic selection. The number of CPS children in each age-race-sex group were adjusted to consistency with census projections by randomly selecting children to be eliminated from the sample households, subject to the constraint that the proportion of childless families be held constant between the base year and the simulation year. This constraint appears to coincide with the historical aversion to childlessness, even as birthrates have declined.

In the economic ageing, income multipliers consistent with CBO's productivity and price increase assumptions were used to inflate non-transfer income. Certain transfer income sources (Social Security, civil service and military pensions, veterans benefits, and Medicare) were imputed to households on the basis of agency estimates as to program costs and caseloads and the distribution of reporters evidenced in the base year file. The imputation technique is briefly described as follows. The probability of a family with given characteristics being selected for imputation is determined by a prior comparison of tabulations of reporters and non-reporters in the relevant dimensions to determine a sampling ratio for each class of beneficiaries. "Imputees" are then randomly selected from among the non-reporters and, employing the Bureau of the Census "hot deck" imputation procedure, they are assigned the amount for the imputed benefit recorded in the last encountered reporter record with the same controlling characteristics (typically age, income from other sources and for family composition). Aggregated household benefits are then summed, compared with the exogenous controls, and benefits are then inflated across the board to meet the control.

THE SIMULATION RESULTS

Program costs and caseloads were simulated on the basis of three different methods for indexing benefit levels and eligibility schedules: Current law, indexing of all programs for price inflation, and indexing of all programs for changes in productivity as well as in the price level. In addition, current law indexing features for social security and military and civil service retirement were varied among the three cases. Table I provides a summary of adjustments made to benefit levels for each program for each of the three cases simulated. Simulation of federal income and payroll taxes and the costs and caseloads for income-tested programs (including Supplemental Security Income, Aid to Families with Dependent Children and Food Stamps) were conducted using the Micro-Analysis of Transfers to Households (MATH) model, a modified version of the Transfer Income Model (TRIM). MATH was also employed to simulate the eligible population for Medicaid, to which the average insurance value of Medicaid benefits was imputed.

The major findings of the study are easily summarized. As shown in Table II, federal transfer program costs, measured in current dollars, are projected to increase substantially by the year 2000. Total costs, compared to their 1975 level increased by 142% under case 1 and 167% under case 3. By comparison over this period, population is projected to increase by 21% and aggregate real non-transfer income by 156%. The projected increase is, however, far lower than that which would be produced by a continuation of the trend experienced between 1955 and 1975. The significant dampening of this trend is unsurprising since the previous period witnessed several phenomena which are not anticipated in the year 2000 projections - several major new programs were introduced (including Medicare, Medicaid, Food Stamps and Supplemental Security Income); benefits in other programs including Social Security and Aid to Families with Dependent Children, were significantly liberalized; and participation rates among eligible populations, previously at relatively low levels in income-tested programs, rose sharply. In addition, the dependent population including children and the retired aged is not projected to grow as rapidly in the next quarter century as in the last 20 years.

Since the absolute cost of transfer programs is not directly meaningful in abstraction from the productive capacity available to support it, aggregate program costs are also presented as a percent of Gross National Product (GNP) in the base year 1975 and the projected year 2000 on the assumption of a 5 percent unemployment rate in both years. Fiscal year 1975 GNP with 5 percent unemployment was estimated by CBO by assuming that a one percentage point change in the unemployment rate would lead to a 3 percent change in GNP. The year 2000 GNP was computed by assuming that GNP in the year 2000 will bear the same relationship to pre-tax/pre-transfer income which obtained in the data base year. The average annual GNP growth rate in constant dollars implied by this procedure is about 3.5%. On this assumption, it is shown in Table II that transfer program expenditures under cases 1 and 2 remain virtually constant as a percent of GNP over the projected period at a level of about 9.4% while under the more generous assumptions of case 3 they rise only modestly to a level of 10.4%.

Table III provides further detail of base year costs and year 2000 projected costs and caseloads for each of the cases simulated. As shown in the table, while aggregate costs do not change dramatically across the three cases, substantial shifts occur in the distribution of benefits and caseloads among programs. For example, between cases 1 and 2, a $19 billion reduction in retirement program costs caused by eliminating the over-compensatory indexing features of social security and government retirement programs is offset by rises in unemployment insurance and by the additional assumptions of automatic cost-of-living increases in all transfer programs.

Several important shifts in the distribution of benefits are noted in the Congressional Budget Office report from which the following three paragraphs are excerpted.[3]

Changes in the Relative Size of Contributory and Noncontributory Programs

In fiscal year 1975, contributory programs - social security and railroad retirement, civil service and military retirement, unemployment insurance and medicare - represented 73 percent of the total cost of Federal income assistance programs. When contributory program costs are estimated using the detailed demographic and economic projections, they represent between 78 and 89 percent of total income assistance expenditures in the year 2000. Generally, the share of the total devoted to programs aimed primarily for the low-income population diminishes.

Program Interactions

In the case of some Federal income assistance programs, the income received from other programs is considered when determining benefits. Such interaction occurs in eligibility and benefit determination for such programs as SSI, AFDC, and food stamps. These program interactions explain part of the cost difference among the various cases presented. In other words, the differences in program costs under the three assumptions cannot be accounted for by price and income adjustments alone.

The food stamp program provides an example of how such interactions work. Between case

1 and 2, food stamp caseloads increased slightly, but total costs actually declined by 21 percent. Since the assumptions regarding the food stamp program were identical for these two cases, another factor - specifically the treatment of other forms of assistance - caused this decline. In case 1, AFDC eligibility and benefit schedules were not adjusted for cost-of-living increases. In case 2, where they were adjusted, AFDC caseloads and benefit levels rose substantially. The net effect of this change in the AFDC program was to reduce food stamp benefits while increasing participation only slightly. That is, because AFDC income is considered in determining food stamp eligibility and benefit levels, an increase in AFDC benefits will decrease food stamp benefits for AFDC recipients. On the other hand, from case 1 to case 2, the removal of social security "coupling" and the retirement "add-on" tended to reduce income for beneficiaries of those programs so that their food stamp benefits rose. On balance, however, the AFDC changes outweighed the effects of the retirement programs, and produced an overall decline in food stamp costs.

Other Outcomes

The detailed nature of the method using demographic and economic projections allowed the effects of many basic assumptions to be isolated. Some of the more important are:

- If participation in all programs increases from the assumed levels to 100 percent, then total income assistance costs in the year 2000 would rise by roughly 1 to 2.5 percent.

- Social security and railroad retirement costs in the year 2000 declined by 7 percent when the "coupling" provision was eliminated.

- Civil service and military retirement costs declined by 11 percent when the 1 percent "add-on" was removed.

- Supplemental Security Income increased dramatically when eligibility and benefit levels were adjusted for increases in percapita income (case 3), because a high proportion of the aged have relatively low incomes.

- AFDC costs would more than triple if a cost-of-living adjustment, not present under current law, were added. However, much of this increase is likely to occur in any case if states continue to raise benefit levels to compensate, at least partially, for inflation induced benefit erosion.

- Because the projected rates of inflation for hospital costs are higher than those assumed for either the general price level or wage increases, Medicare costs were projected to rise rapidly from 1975 to the year 2000.

- Medicaid costs were projected to increase rapidly for the same reason as Medicare costs but, in addition, program interaction raised caseloads, increasing costs in all three cases.

- Increases in real income and in benefit levels from other assistance programs tend to dampen the increase in food stamp costs.

CONCLUSION

Given the very limited time available for the study, it is not difficult to suggest possible improvements and extensions. While the precision obtained in such projections is obviously limited by our ability to foresee the future, it would be desirable to test the sensitivity of the results to certain basic assumptions, most notably, the assumed real income growth rate of 2.6 percent which affects both estimates of eligibility and of GNP. The assumed rate of inflation, 5 percent, is of less concern except in case 1, since it produces offsetting increases in nontransfer income and benefits for each assistance program. However, the Medicaid and Medicare projections are obviously highly sensitive to the assumptions made with regard to all components of the medical price index. It may also be interesting to test the sensitivity of the findings to the implicit assumption of constant age/sex/race labor force participation rates. Assumptions with regard to the age distribution of social security and military and civil service retirement benefits and their concomitant receipt by individuals and households also warrant further inspection and a more adequate treatment of private and state and local government pensions would be desirable. However, given the magnitude of the benefits accounted for, the offsetting nature of the interactions among government transfer programs, and the conservative assumptions employed in accounting for these factors whenever the direction of bias was determinable, it does not seem likely that the overall impact of these latter improvements on the summary results would be significant.

FOOTNOTES

1. U.S. Congress, Senate, Committee on the Budget, Growth of Government Spending for Income Assistance: A Matter of Choice, Prepared by the Congressional Budget Office, 94th Congress, 1st session, 1975.

2. For a detailed technical description of the procedures employed in this study see Jodie T. Allen and Raymond J. Uhalde, Long-Range Estimates of the Costs and Caseloads of the Major Income Assistance Programs, Project Report Series, no. 76-04 (Washington, D. C.: Mathematica Policy Research, October 1975).

3. Senate Committee on the Budget, Growth of Government Spending for Income Assistance, pp. 10-12.

TABLE I

ASSUMPTIONS ABOUT COST OF LIVING AND REAL INCOME ADJUSTMENTS
FOR EACH PROJECTION CASE, BY GOVERNMENT PROGRAM

Program	Current Law (1)	Adjusted for cost of living (2)	Adjusted for income (3)
Social Security and Railroad Retirement	Benefits doubly indexed for cost of living. Situation known as "coupling."	Benefits indexed for cost of living. Benefits formula "decoupled."	Benefits indexed for cost of living. Benefits formula "decoupled."
Civil Service and Military Retirement.	Benefits indexed for cost of living plus 1 percent. Situation known as 1 percent "add-on."	Benefits indexed for cost of living. One percent "add-on eliminated.	Benefits indexed for cost of living. One percent "add-on" eliminated.
Unemployment Insurance.	In some States, indexed for cost of living and real wages (because maximum benefit is a function of average State wage). In other States, benefits not indexed.	In some States, indexed for cost of living plus real income (because maximum benefit is a function of average State wage). In other States, benefits indexed for cost of living.	Benefits indexed for cost of living and real incomes in all States.
Veterans Benefits.	Benefits not indexed in law.	Benefits indexed for cost of living.	Benefits indexed for cost of living and real income.
Medicare.	Benefits indexed for projected increases in factor costs and services provided.	Same as case 1.	Same as case 1.
Medicaid.	Same as medicare.	Same as medicare.	Same as medicare.
Supplemental Security Income. Participation Rate: 90 percent.	Federal benefits indexed for cost of living. State benefits assumed to disappear by year 2000.	Benefits indexed for cost of living.	Benefits indexed for cost of living and real wages.
Aid to Families with Dependent Children Participation Rate: 90 percent.	Benefits not indexed in law.	Benefits indexed for cost of living.	Benefits indexed for cost of living and real income.
AFDC-Unemployed Fathers. Participation Rate: 50 percent.	Same as AFDC	Same as AFDC.	Same as AFDC
Food Stamps.	Benefits indexed for cost of living.	Benefits indexed for cost of living.	Benefits indexed for cost of living and real income.
Participation Rate:*			
Non-Public Assistance	48 percent	48 percent	46 percent
Public Assistance	93 percent	84 percent	70 percent

*NOTE: Demographic changes in conjunction with participation rates which differ by benefit level and family size result in different participation rates for each of the program cases.

SOURCE: Senate Committee on the Budget, Growth of Government Spending for Income Assistance.

TABLE II

EXPENDITURES ON FEDERAL INCOME ASSISTANCE PROGRAMS FOR FISCAL YEARS 1955 AND 1975, AND ALTERNATIVE PROJECTIONS[1]

(Constant 1975 dollars - in billions)

	Fiscal year		Year 2000			
	1955 (actual)	1975 (esti-mate)	Recent trend[2]	Current law (1)	Adjusted for cost of living (2)	Adjusted for income (3)
Total cost	25.7	142.9	1,220.1	345.1	341.8	381.6
Contributory programs	15.1	103.8	886.3	307.8	293.5	297.4
Noncontributory programs	10.6	39.1	333.8	37.3	48.3	84.2
Total cost as percent of GNP[3]	3.4	9.3	33.4	9.4	9.4	10.4

[1]Expenditure figures include State and local government portion of Medicaid and AFDC. GNP and unemployment insurance have been adjusted to the levels that would exist under a 5-percent unemployment rate.

[2]The 1955 to 1975 growth rate was applied to total costs. The contributory and noncontributory shares of total costs were assumed to be the same as in 1975.

[3]GNP figures at 5 percent unemployment for fiscal years 1955, 1975, and 2000 are $755 billion, $1,529 billion, and $3,653 billion, respectively. GNP for fiscal 1975 with 7.4 percent unemployment is estimated at $1,426 billion.

SOURCE: Same as Table I. Base numbers derived from subsequent chapters of this report.

TABLE III

CASELOADS AND COSTS OF FEDERAL INCOME ASSISTANCE PROGRAMS

(Constant 1975 dollars - in billions)

	Fiscal year 1975[1]		Year 2000: Case 1 - Current law			Year 2000: Case 2 - Adjusted for cost of living			Year 2000: Case 3 - Adjusted for income		
	Costs[2]	Costs as percent of GNP[3]	Family case-load (thousands)	Costs[2]	Costs as percent of GNP[3]	Family case-load (thousands)	Costs[2]	Costs as percent of GNP[3]	Family case load (thousands)	Costs[2]	Costs as percent of GNP[3]
Social Security and Railroad Retirement	66.7	4.4	31,411	178.9	4.9	31,411	166.2	4.5	31,411	166.2	4.5
Civil Service and Military Retirement	13.2	0.9	3,642	53.5	1.5	3,642	47.5	1.3	3,642	47.5	1.3
Unemployment Insurance	[4] 9.1	0.6	5,415	11.7	0.3	5,415	16.0	0.4	5,415	19.9	0.5
Medicare - Part A	10.6	0.7	[5] 33,180	48.6	1.3	[5] 33,180	48.6	1.3	[5] 33,180	48.6	1.3
Part B	4.2	0.3	[5] 33,273	15.2	0.4	[5] 33,273	15.2	0.4	[5] 33,273	15.2	0.4
Subtotal: Contributory programs	103.8	6.8	([10])	307.9	8.4	([10])	293.5	7.9	([10])	297.4	8.0
Veterans Benefits	7.6	0.5	3,604	1.6	[6] 0.0	3,604	5.3	0.1	3,604	9.7	0.3
Supplemental Security Income	4.8	0.3	1,405	1.9	0.1	1,614	2.0	0.1	4,535	7.2	0.2
AFDC	9.7	0.6	3,288	2.3	0.1	3,815	8.8	0.2	4,466	18.7	0.5
Medicaid	12.4	0.8	[5] 29,799	24.0	0.7	[5] 31,713	26.3	0.7	[5] 52,891	41.5	1.1
Food Stamps	[7] 4.6	0.3	[8] 6,800	7.5	0.2	[8] 6,885	5.9	0.2	[8] 10,814	7.1	0.2
Subtotal: Noncontributory programs	39.1	2.6	([10])	37.3	1.1	([10])	48.3	1.3	([10])	84.2	2.3
Total: All programs[9]	142.9	9.3	([10])	345.1	9.4	([10])	341.8	9.4	([10])	381.6	10.4

[1] Caseloads not available on a comparable basis for fiscal year 1975.

[2] Includes administrative expenses.

[3] GNP for fiscal year 1975 with 5 percent unemployed is estimated to be $1,529 billion; year 2000 GNP in 1975 dollars is estimated to be $3,653 billion.

[4] Adjusted to a 5 percent rate of unemployment. The actual unemployment insurance for fiscal year 1975 was estimated at $13.5 billion with a 7.4 percent unemployment rate.

[5] Eligible persons.

[6] Zero due to rounding.

[7] Includes commodity distribution.

[8] Caseloads are average monthly households.

[9] Components may not sum to total due to rounding.

[10] Cannot be summed because of multiple program beneficiaries.

SOURCE: Same as Table I. Derived from "Current Service Estimates for Fiscal Year 1977," OMB, November 10, 1975; "Social Security Bulletin," February 1975; and special tabulations prepared by Mathematica, Inc. (see Chapters I and II).

ALBANY (NY) HEALTH REGION FAMILY PLANNING SURVEY: FACTORS UNDERLYING RELIGIOUS AND SOCIOECONOMIC DIFFERENCES IN PERIOD FERTILITY

John E. Anderson, Center for Disease Control

INTRODUCTION

This paper deals with the measurement of fertility differences between various subgroups of women, using data from the 1974 Albany (NY) Health Region Family Planning Survey; in particular, fertility differences between women residing in high and low socioeconomic areas, of different educational levels, and of different religions are discussed. The effects of 3 basic factors on fertility differences are measured: marital fertility, out-of-wedlock fertility, and the distribution of women by marital status and marriage duration.

Briefly, the Albany (NY) Health Region Family Planning Survey was a household probability sample of all women 15-44 years of age living in the 18-county Albany Health Region (AHR) of New York State, an area that includes a large metropolitan area (Albany -Troy -Schenectady SMSA), and rural areas. The survey of 2,059 women was conducted in the summer of 1974 by the New York State Department of Health, with consultation from the Family Planning Evaluation Division, Center for Disease Control, Atlanta, Georgia (1).

While this survey had a primarily public health and family planning orientation, it was designed to yield measures of fertility--total, planned, and unplanned--comparable to those obtained from national surveys such as the National Fertility Studies of 1965 and 1970. National surveys of this kind have been done regularly for over 20 years and have documented social variation in fertility and how it has changed. For example, Ryder has described the convergence in marital fertility over the 1960s as measured by the 2 National Fertility Studies as follows:

> In summary, the differences between religious groups like those between the races and educational levels, have diminished appreciably during the past decade for wanted fertility, for unwanted fertility, and for modes of fertility regulation (2).

The surveys to which Ryder refers and, in fact, all national surveys done so far, have essentially included only ever-married women. (The 1973 and 1976 National Surveys of Family Growth also include never-married mothers living with their own children, but not other never-married women.) These studies, then, only measure marital fertility. While falling within the tradition of these national surveys, the AHR study differs in the important respect that all women 15-44, regardless of marital status, were included. Thus, the fertility of all women can be measured, and the separate effects of marital fertility, marital status and duration, and out-of-wedlock fertility can be examined. While marital fertility is by far the largest component of overall fertility levels, other factors can be important in producing fertility differences between various groups. Differences in marital fertility, as described by national survey data, can be of different magnitude or even of a different direction than differences in overall fertility. Data from the AHR survey show how the factors of marital fertility, out-of-wedlock fertility, and marital status combine to produce the observed social differences in fertility.

Fertility Measurement

The fertility measure used as a dependent variable in this analysis is the number of births per woman in the interval between January 1, 1969 and the survey date (Summer, 1974). Current pregnancies were counted as live births so the interval covered was approximately 6 years. This fertility measure is analogous to a general fertility rate aggregated over 6 years.

Each woman interviewed in the survey was asked a series of detailed questions about each pregnancy she had had in the interval. Included were a sequence of questions (comparable to the 1970 National Fertility Study Questionnaire) about whether the woman had wanted to become pregnant at the time of conception and if she had not, whether she had wanted a child later, or had wanted no more children. On the basis of these questions, each live birth the woman had had since 1969 was classified as "planned," "mistimed," or "unwanted." In addition, 3% of the births were classified as unknown because of insufficient data.

Planned births were defined as those that were desired and did not occur before planned; mistimed are those that occurred before planned but were still desired, and unwanted are "number failures" or those that were in excess of the desired number. These 4 categories of births (including unknown) are mutually exclusive and exhaustive. In addition, the categories of planned and mistimed can be combined into "wanted births" while mistimed and unwanted form the category "unplanned births." This usage conforms to that of published analyses of the 1965 and 1970 National Fertility Studies.

Over the 6-year period, these figures indicate that there were 0.46 births per woman. About 54% of these births were planned, 29% were mistimed, and 14% were unwanted. How do these figures compare with national estimates? In order to make this comparison, the data must be limited momentarily to married women and adjusted to represent a 5- rather than 6-year interval. When these adjustments are made, total births per married woman in the AHR survey (Table 2) is 0.56, compared with 0.68 for the period 1966 to 1970 in the 1970 National Fertility Study. The lower fertility in AHR is not unexpected in view of recent trends in fertility. Despite this overall fertility difference, Table 1 indicates that the percent distribution of births by planning status in AHR is very similar to that found in the national studies.

Differences in Number of Births per Woman

Table 2 summarizes the pattern of differences in fertility found in the AHR survey, which is reported in greater detail elsewhere(3).

Fertility was lowest among non-high school graduates, but this reflects the large number of teenage unmarried high school students in this category. If the rates are standardized for marital status and duration, a different pattern appears(Table 2). Holding these factors constant, fertility has an inverse relationship with education. Percent of births that are unwanted--in excess of the desired number--also becomes higher with less education.

Fertility is higher in lower status women regardless of whether the rates are standardized (Table 2). (The index of socioeconomic status used here is that developed by the New York State Department of Health and is based on 1970 census data. Each census tract within the SMSA and each minor civil division outside the SMSA were assigned to a socioeconomic category, based on median education, percent of labor force in unskilled occupations, and median income.) The percent unwanted is higher in women of lower socioeconomic status, indicating less control over fertility.

Finally, Table 2 shows that among white women fertility was lower for Catholics than for non-Catholics. As discussed later, the difference in percent married and the marriage duration distribution are important factors. When these are held constant through standardization, there is no difference in fertility by religious affiliation and little difference in the percent of births unwanted.

One important variable in fertility difference in the U.S.--race--is not examined here because the AHR is predominantly white, and too few black women were interviewed for separate analysis. It is necessary to control for race, however, in some cases, since blacks tend to be concentrated in the lower status categories.

Underlying Factors

Table 3 shows the distribution of women by marriage duration, socioeconomic area, and religion. Women of low socioeconomic status are more concentrated in the never-married category (where fertility is low) and in the most recently married category (where fertility is highest) than women of higher status. This pattern is effected only slightly by confining the analysis to whites. Compared with other women, Catholics are more likely to be never-married or married more than 15+ years, which works against high fertility.

Table 4 shows births per woman by category of marriage duration. Out-of-wedlock fertility is considerably higher among low socioeconomic status women. The effect of high black out-of-wedlock fertility is shown in that there are .19 births per never-married woman in the low socioeconomic groups for all races; this figure declines to .10 births per woman when looking at whites only, but is still more than twice as high as in the higher status woman. Fertility is higher for women in the low socioeconomic strata married either less than 5 years or 15 or more years, consistent with the finding that these women were less successful in both spacing and limiting births. By religion, non-Catholics had twice as high out-of-wedlock fertility, Catholics had substantially higher fertility for those recently married, and there is no consistent pattern of difference for women married 5 years or more. This conforms to earlier findings from the New York Survey and the National Fertility Study, which show that Catholics are less efficient spacers, but are as good as non-Catholics in limiting the number of births (4).

Table 5 illustrates how the factors just discussed--marriage duration, fertility within categories of marriage duration, and out-of-wedlock fertility--worked to bring about the observed difference in births per woman between the woman living in low socioeconomic status areas and higher status areas. The method used, components of difference, standardizes for each of these 3 factors separately and assigns a portion of the difference in births per woman to each factor based on this standardization. An interaction term is necessary because some of the difference cannot be accounted for by changing these factors separately, but is due to the fact that they vary together (5). If the socioeconomic differences are analyzed in this way for all races, out-of-wedlock fertility accounts for the largest share of the difference (49%), followed by marital fertility, and marriage duration distribution. All 3 of these components are positive, indicating that the factors all contribute toward higher fertility among women of low socioeconomic status. The large effect for out-of-wedlock fertility reflects the pattern of childbearing of the black women, who make up 30% of all women in the low socioeconomic status. In the AHR, as in the United States, blacks differ significantly from whites in having a relatively large share of their childbearing in the out-of-wedlock category (6). When attention is restricted to white women, the total difference in fertility is about the same, but the rank-ordering of the effects changes. Marital fertility is the most important factor in the socioeconomic difference in overall fertility for white women, followed by the marriage duration distribution, and out-of-wedlock fertility, which still accounts for about one-fifth of the difference.

In the AHR survey, then, the socioeconomic difference in births per woman results from a number of factors working in concert--the distribution of women by marital status and marriage duration, the level of out-of-wedlock fertility, and the rates of fertility for women within categories of marriage duration. The relative importance of these factors depends on which ethnic group is examined, but in any case, the factors all contribute to the observed pattern of difference.

In contrast, the effects are in 2 directions for religious difference shown in Table 6. Fertility is higher for non-Catholics (Table 6). The difference in marriage duration alone could account for this difference. Out-of-wedlock fertility also works to depress Catholic fertility relative to non-Catholic. Only marital fertility within marriage duration categories works on the average toward higher Catholic fertility.

Conclusion

These findings illustrate the value of interviewing all women 15-44 years of age regardless of marital status. The kinds of socioeconomic differences in fertility observed in the survey indicate a need for public effort to combat inequities in fertility control, even at a time when fertility levels are very low. If this is really worth doing, the dynamics involved in the observed fertility differences need to be measured. This requires designing comprehensive national surveys that measure fertility of all segments in the population.

REFERENCES

1. Morris L, Conn J, Stroh G, and Gesche M: "Sexual Experience, Contraceptive Usage, and Source of Contraception for Never-Married Women: Albany (NY) Health Region," presented at the Annual Meeting of the Population Association of America, Montreal, Canada, April 1976

2. Ryder NB: "Recent Trends and Group Differences in Fertility," in CF Westoff, ed., Toward the End of Growth: Population in America, Prentice Hall, Englewood Cliffs, NJ, 1973, p 65

3. Anderson JE, Morris L, and Gesche M: "Planned and Unplanned Fertility in Upstate New York, 1969-1974," Family Planning Perspectives, 1976, forthcoming

4. Ibid; Ryder NB: "Contraceptive Failure in the United States," Family Planning Perspectives 5, p 133, 1973

5. Kitagawa EM: "Components of Difference Between Two Rates," Journal of the American Statistical Association, Vol 50, p 1168, 1955

6. Presser HB: "The Timing of the First Birth, Female Roles, and Black Fertility," The Millbank Memorial Fund Quarterly, Vol 49, p 335, 1970; Rindfuss R and Westoff CF: "The Initiation of Contraception," Demography, Vol II, p 75, 1975; Anderson JE and Smith JC: "Planned and Unplanned Fertility in a Metropolitan Area: Black and White Differences," Family Planning Perspectives Vol 7, p 281, 1976

TABLE 1

Births Per Married Woman and Percent of Births in the Period 1969-1974, Albany (NY) Health Region Family Planning Survey and Comparable Data for 1966-1970 from the 1970 National Fertility Study (NFS)

Planning Status	Births per Married Woman: Albany		Births per Married Woman: NFS*	Percent: Albany	Percent: NFS**
TOTAL	.67		.68	*100*	*100*
Planned	.37	.55	.58	*56*	*57*
Mistimed	.18			*27*	*29*
Unwanted	.10		.10	*14*	*14*
Unknown	.02		--	*3*	--

*Ryder NB: "Recent Trends and Group Differences in Fertility," in CF Westoff, ed., Toward the End of Growth: Population in America, Prentice Hall, Englewood Cliffs, NJ, 1973, p 63, Table 6-1

**Westoff CF: "The Decline of Unplanned Births in the United States," Science 191, January 9, 1976, p 38

TABLE 2

Births per Woman 15-44 Unstandardized and Standardized, and Percent of Births Unwanted 1969-1974, by Education, Socioeconomic Area and Religion Albany (NY) Health Region Family Planning Survey

		Education			Socio-economic Area		Religion (white only)	
Unstandardized	Total	College	High Schl Grad	Less than High Schl Grad	Lower Middle & Above	Low	Catholic	Non-Catholic
Births per Woman	.46	.51	.50	.35	.45	.55	.44	.49
Percent Unwanted	*14.9*	*12.3*	*14.9*	*18.5*	*14.5*	*17.5*	*15.4*	*14.9*
Standardized*								
Births per Woman	.46	.40	.47	.55	.45	.54	.46	.46
Percent Unwanted*	*14.9*	*13.5*	*13.5*	*17.3*	*13.8*	*18.9*	*14.6*	*14.5*
No. of Women (unweighted)	2059	559	896	600	1427	632	988	877
No. of Births (unweighted)	1161	317	528	312	759	402	568	493

*Standardized for marriage duration

TABLE 3

Proportion of Women by Marital Duration by Socioeconomic Area and Religion Albany (NY) Health Region Family Planning Survey

		Socioeconomic Area				Religion (white only)	
		All Races		White Only			
Marital Duration	Total	Lower Middle & Above	Low	Lower Middle & Above	Low	Catholic	Non-Catholic
Never Married	*.31*	*.30*	*.35*	*.30*	*.33*	*.32*	*.28*
<5 Years	*.18*	*.17*	*.21*	*.18*	*.24*	*.17*	*.19*
5-9 Years	*.15*	*.16*	*.14*	*.15*	*.14*	*.14*	*.17*
10-14 Years	*.14*	*.14*	*.11*	*.15*	*.11*	*.13*	*.15*
15+ Years	*.21*	*.22*	*.18*	*.22*	*.18*	*.22*	*.20*
Unknown	*.01*	*.01*	*.01*	*.01*	*.00*	*.01*	*.00*
TOTAL	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*

TABLE 4

Births per Woman 15-44, 1969-1974, by Marriage Duration by Socioeconomic Area and Religion Albany (NY) Health Region Family Planning Survey

		Socioeconomic Area				Religion (White Only)	
		All Races		White Only			
Marriage Duration	Total	Lower Middle & Above	Low	Lower Middle & Above	Low	Catholic	Non-Catholic
Never Married	*.06*	*.04*	*.19*	*.04*	*.10*	*.03*	*.06*
<5 Years	*1.14*	*1.13*	*1.22*	*1.15*	*1.24*	*1.23*	*1.08*
5-9 Years	*.99*	*1.00*	*.92*	*1.02*	*.99*	*1.00*	*1.04*
10-14 Years	*.42*	*.42*	*.43*	*.43*	*.46*	*.38*	*.48*
15+ Years	*.10*	*.09*	*.23*	*.10*	*.22*	*.11*	*.10*
Unknown	*.45*	*.45*	*1.67*	*.29*	*1.00*	*.43*	*.00*
TOTAL	*.46*	*.45*	*.55*	*.46*	*.56*	*.44*	*.49*

TABLE 5

Components of Difference, Births per Woman, 1969-1974, for Socioeconomic Area All Races and White Only Albany (NY) Health Region Family Planning Survey

Births per Woman	All Races	Percent	Whites Only	Percent
Low Socioeconomic Area	.55		.56	
Lower Middle & Above	.45		.46	
Difference	.10	100.0	.10	100.0
Component due to:				
Marriage Duration	.02	17.0	.03	32.0
Marital Fertility	.04	41.0	.04	46.0
Out-of-Wedlock Fertility	.05	49.0	.02	21.0
Interaction	-.01	- 7.0	.00	1.0

TABLE 6

Components of Difference, Births per Woman, 1969-1974 for Catholics and Non-Catholics (White Only) Albany (NY) Health Region Family Planning Survey

Births per Woman		Percent
Catholics	.44	
Non-Catholics	.49	
Difference	-.05	100.0
Component due to:		
Marriage Duration*	-.05	104.0
Marital Fertility	.01	- 25.0
Out-of-wedlock Fertility	-.01	18.0
Interaction	-.00	2.0

*Note: Negative component contributes toward lower fertility among Catholics

ON THE EFFICIENCY COMPARISON OF TWO πps SAMPLING STRATEGIES

C. Asok and B. V. Sukhatme, Iowa State University

1. Introduction: Let $\{U_1, U_2, \ldots, U_N\}$ be a finite population of a known number N of identifiable units and consider the problem of estimating the population total Y of a characteristic y with the value Y_i on unit U_i. When information on an auxiliary characteristic x highly correlated with y is available it is often advantageous to select a sample of size n with varying probabilities and without replacement. For any sampling design an unbiased estimator of Y proposed by Harvitz and Thompson is

$$\hat{Y}_{H.T.} = \sum_{i \epsilon s} Y_i/\pi_i, \tag{1.1}$$

where the sum is over all distinct units of the sample s and π_i is the probability of including U_i in a sample of size n. The variance of $\hat{Y}_{H.T.}$ is given by

$$V(\hat{Y}_{H.T.}) = \sum^{N} \frac{Y_i^2}{\pi_i} + \sum_{i}^{N} \sum_{j(\neq i)}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} Y_i Y_j - Y^2, \tag{1.2}$$

where π_{ij} is the probability for the i-th and j-th units to be both in the sample. (1.2) evidently reduces to zero when π_i is proportional to Y_i which suggests that considerable reduction in the variance can be achieved by making $\pi_i \propto x_i$. Such a scheme must obviously satisfy the condition

$$\pi_i = np_i , \tag{1.3}$$

where $p_i = X_i/X$, X being the sum of all the X_i's.

Several schemes have been proposed in the literature that satisfy condition (1.3). However not many of them are applicable for sample size greater than two. Further none of these procedures, owing to the complications involved, are strictly applicable in large scale surveys. In this connection it is worthwhile to quote Durbin (1953, p. 267). He says:

"The strict application of the usual methods of unequal probability sampling without replacement, including the calculation of unbiased estimates of sampling error is out of the question in certain kinds of large-scale survey work on grounds of practicability. There is therefore a need for methods which retain the advantages of unequal probability sampling without replacement but are rather easier to apply in practice and only involve a slight loss of exactness." Also, practically nothing is known as to how the different procedures compare among themselves as measured by the variance of the corresponding estimators. In an earlier article the authors have compared the procedure of Goodman and Kish with that of Sampford and have concluded that Sampford's procedure yields a uniformly better estimator than the procedure of Goodman and Kish. In this paper we compare the procedure of Goodman and Kish with that of Hanurav. Also we confine to the simple but important case of sample size 2. We desecribe the Hanurav's procedure in the following:

Without loss of generality let

$$0 < X_1 \le X_2 \le \cdots \le X_{N-1} \le X_N \le \frac{X}{2} . \tag{1.4}$$

we first describe sampling scheme A covering the special case

$$X_{N-1} = X_N . \tag{1.5}$$

Sampling Scheme A:

Step 1. Select two units from the population with probability p_K for U_K and with replacement. If the sample consists of distinct units accept it; otherwise reject the sample and proceed to Step 2. Select two units from the population with probabilities proportional to p_K^2 and with replacement. Again, if the sample consists of distinct units accept it; otherwise reject and proceed to further steps. In general, if the 1st, 2nd, ... (m-1)th steps result in rejections, the units are drawn in the m^{th} step with probabilities proportional to $p_1^{2^{m-1}}, p_2^{2^{m-1}}, \ldots p_N^{2^{m-1}}$. It has been shown by Hanurav that sampling scheme A terminates after a finite number of steps with probability 1. Also the inclusion probabilities π_i and π_{ij} are given by

$$\pi_i = 2p_i \tag{1.6}$$

and
$$\pi_{ij} = 2p_i p_j [1 + \sum_{K=1}^{\infty} W_K] , \tag{1.7}$$

where
$$W_K = \frac{(p_i p_j)^{2^K - 1}}{S_{(1)} S_{(2)} \cdots S_{(K)}} \tag{1.8}$$

with
$$S_{(t)} = \sum_{K=1}^{N} p_K^{2^t} . \tag{1.9}$$

Now restriction (1.5) is dropped to generalize scheme A. This generalized scheme, which is denoted as sampling scheme B, is described as follows:

Sampling Scheme B:

Step 1. Conduct a binomial trail with probability of success δ given by

$$\delta = \frac{2(1-p_N)(p_N - p_{N-1})}{(1-p_N-p_{N-1})} . \tag{1.10}$$

If the trial results in success proceed to step 2; otherwise proceed to step 3.

Step 2. Select one of the units $U_1, U_2, \ldots, U_{N-1}$ with probabilities proportional to $p_1, p_2, \ldots, p_{N-1}$. If U_j is the unit thus selected, accept U_N and U_j as the unordered sample.

Step 3. Proceed with the sampling scheme A with the probabilities p_i replaced by p_i^* given by

$$p_i^* = \frac{p_i}{1-p_N+p_{N-1}} \quad \text{for} \quad 1 \le i \le N-1$$

$$\text{and} \quad p_N^* = p_{N-1}^* = \frac{p_N - \frac{1}{2}\delta}{1-\delta} \qquad (1.11)$$

For sampling scheme B, π_i and π_{ij} are given by

$$\pi_i = 2p_i \qquad (1.12)$$

$$\pi_{ij} = (1-\delta)\phi_{ij} \text{ for } 1 \le i \ne j < N,$$

$$\text{and} \quad \pi_{Nj} = \delta\{p_j/(1-p_N)\} + (1-\delta)\phi_{Nj}, \qquad (1.13)$$

$$\text{where} \quad \phi_{ij} = 2p_i^*p_j^* \, (1 + \sum_{K=1}^{\infty} W_K^*), \qquad (1.14)$$

$$W_K^* = \frac{(p_i^*p_j^*)^{2^K-1}}{S_{(1)}^* S_{(2)}^* \cdots S_{(K)}^*}, \qquad (1.15)$$

$$S_{(t)}^* = \sum_{K=1}^{N} p_K^{*2^t}. \qquad (1.16)$$

Contrary to Hanurav's claim it can be seen from (1.7) and (1.13) as to how complicated it is to calculate the pairwise probabilities. Since the expression for pairwise probability is in terms of an infinite series, its exact numerical value for given data can never be calculated and as such one must resort to some kind of approximation for getting the pairwise probabilities and hence the variance. As the condition $\pi_{ij} < \pi_i\pi_j$ is satisfied for this procedure one can conclude that it yields an estimator which has a uniformly smaller variance than the customary estimator in sampling with replacement. As the method has been subsequently extended by Hanurav to cover the case of arbitrary sample size it will be of interest to study the relative performance of this method relative to the procedure of Goodman and Kish and of Sampford. Hartley and Rao used an asymptotic approach for deriving the expression for π_{ij} of the Goodman and Kish procedure and hence the variance of the H.T. estimator. As such it would be realistic for comparison purposes to derive the approximate expressions for π_{ij} and hence the variance for the Hanurav's procedure using the asymptotic approach of Hartley and Rao. These approximations should be of value for their own sake, since the simplicity of computation is one of the factors to be considered in choosing a sampling procedure. We will first evaluate the π_{ij} and hence the variance for scheme A under the asumption of Hartley and Rao viz., N is large and p_i is of $O(N^{-1})$.

2. Evaluation of π_{ij} and $V(\hat{Y}_{H.T.})$ for scheme A:

In order to evaluate the variance correct to $O(N^1)$ we have to evaluate π_{ij} correct to $O(N^{-3})$. Also for using in the case of smaller size populations the variance correct to $O(N^0)$ is to be evaluated by evaluating π_{ij} correct to $O(N^{-4})$.

When p_i is of $O(N^{-1})$ it can be easily seen that $S_{(t)}$ is of $O(N^{-2^t+1})$ from which it follows that W_K will be of $O(N^{-K})$.

Hence the expression for π_{ij} correct to $O(N^{-4})$ is

$$\pi_{ij} = 2p_ip_j[1 + \frac{p_ip_j}{\Sigma p_t^2} + \frac{p_i^3p_j^3}{\Sigma p_t^2 \Sigma p_t^4}]. \qquad (2.1)$$

Substituting from (1.6) and (2.1) into (1.2), simplifying and retaining terms to $O(N^0)$ only we get

$$V(\hat{Y}_{H.T.})_A = \Sigma \frac{Y_i^2}{2p_i} + \frac{1}{2}[Y^2 - \Sigma Y_t^2] + \frac{1}{2\Sigma p_t^2}[(\Sigma Y_t p_t)^2 - \Sigma Y_t^2 p_t^2] + \frac{1}{2\Sigma p_t^2 \Sigma p_t^4}(\Sigma Y_t p_t^3)^2 - Y^2$$

$$= \frac{1}{2}[\Sigma \frac{Y_t^2}{p_t} - Y^2] - \frac{1}{2}[\Sigma Y_t^2 - \frac{(\Sigma Y_t p_t)^2}{\Sigma p_t^2}] - \frac{1}{2\Sigma p_t^2}[\Sigma Y_t^2 p_t^2 - \frac{(\Sigma Y_t p_t^3)^2}{\Sigma p_t^4}]$$

$$= \frac{1}{2}\Sigma p_t z_t^2 - \frac{1}{2}[\Sigma p_t^2 z_t^2 - \frac{(\Sigma p_t^2 z_t)^2}{\Sigma p_t^2}] - \frac{1}{2\Sigma p_t^2}[\Sigma p_t^4 z_t^2 - \frac{(\Sigma p_t^4 z_t)^2}{\Sigma p_t^4}] \qquad (2.2)$$

where $z_t = \frac{Y_t}{p_t} - Y$.

When the variance is considered to $O(N^1)$ only we get

$$V(\hat{Y}_{H.T.})_A = \frac{1}{2}\Sigma p_t z_t^2 - \frac{1}{2}[\Sigma p_t^2 z_t^2 - \frac{(\Sigma p_t^2 z_t)^2}{\Sigma p_t^2}] \qquad (2.4)$$

The term of $O(N^2)$ in the above viz., $\frac{1}{2}\Sigma p_t z_t^2$ is the variance of the customary estimator in the case of sampling with replacement. Hence the term of $O(N^1)$, viz., $\frac{1}{2}[\Sigma p_t^2 z_t^2 - \frac{(\Sigma p_t^2 z_t)^2}{\Sigma p_t^2}]$ represents the reduction in variance achieved by adopting scheme A of Hanurav. Following the method adopted in obtaining (2.2) one can in fact obtain the exact variance for scheme A as follows:

From (1.6) and (1.4) with $W_0 = 1$ we get

$$\sum_i \sum_{j(\ne i)} \frac{\pi_{ij}}{\pi_i\pi_j} Y_iY_j = \frac{1}{2}\sum_i \sum_{j(\ne i)} Y_iY_j(\sum_{K=0}^{\infty} W_K),$$

$$= \frac{1}{2}\sum_{K=0}^{\infty}\sum_i^N \sum_{j(\ne i)}^N W_K Y_i Y_j$$

$$= \frac{1}{2}\sum_{K=0}^{\infty} R_K \qquad (2.5)$$

where $R_K = \sum_{i}^{N} \sum_{j(\neq i)}^{N} W_K Y_i Y_j$

$$= \frac{1}{S_{(1)}S_{(2)}\cdots S_{(K)}}[(\Sigma Y_t p_t^{2^K-1})^2 - \Sigma Y_t^2 p_t^{2^{K+1}-2}] .$$

Substituting from (2.5) into (1.2) we get

$$V(\hat{Y}_{H.T.})_A = \Sigma \frac{Y_t^2}{2p_t} - \frac{1}{2}\Sigma Y_t^2 + \frac{1}{2\Sigma p_t^2}[(\Sigma Y_t p_t)^2 - \Sigma Y_t^2 p_t^2] + \frac{1}{2\Sigma p_t^2 \Sigma p_t^4}[(\Sigma Y_t p_t^3)^2 - \Sigma Y_t^2 p_t^6] + \ldots$$

Rearranging the terms we get

$$V(\hat{Y}_{H.T.})_A = \frac{1}{2}[\Sigma \frac{Y_t^2}{p_t} - Y^2] - \sum_{r\geq 1} \frac{1}{2\Sigma p_t^2 \Sigma p_t^4 \ldots \Sigma p_t^{2^{r-1}}}[\Sigma Y_t^2 p_t^{2^r-2} - \frac{(\Sigma Y_t p_t^{2^r-1})^2}{\Sigma p_t^{2^r}}]$$

Substituting $\Sigma z_t^2 p_t^{2^r} - \frac{(\Sigma z_t p_t^{2^r})^2}{\Sigma p_t^{2^r}}$ for $\Sigma Y_t^2 p_t^{2^r-2} - \frac{(\Sigma Y_t p_t^{2^r-1})^2}{\Sigma p_t^{2^r}}$ and $\Sigma p_t z_t^2$ for $\Sigma \frac{Y_t^2}{p_t} - Y^2$ in the above we get

$$V(\hat{Y}_{H.T.})_A = \frac{1}{2}\Sigma p_t z_t^2 - \sum_{r\geq 1} \frac{1}{2\Sigma p_t^2 \Sigma p_t^4 \ldots \Sigma p_t^{2^{r-1}}} [\Sigma p_t^{2^r} z_t^2 - \frac{(\Sigma p_t^{2^r} z_t)^2}{\Sigma p_t^{2^r}}] \qquad (2.6)$$

3. Evaluation of π_{ij} and $V(\hat{Y}_{H.T.})$ for Scheme B:

For evaluating $V(\hat{Y}_{H.T.})_B$ correct to $O(N^0)$ we have to evaluate π_{ij} correct to $O(N^{-4})$.

From (1.11) we get by expanding,

$$p_i^* = \frac{p_i}{1-p_N+p_{N-1}} = p_i\{1 + (p_N-p_{N-1}) + (p_N-p_{N-1})^2 + (p_N-p_{N-1})^3 + \ldots\} \text{ for } i = 1,2,\ldots,N-1$$

and

$$p_N^* = p_{N-1}\{1 + (p_N-p_{N-1}) + (p_N-p_{N-1})^2 + (p_N-p_{N-1})^3 + \ldots\} . \qquad (3.1)$$

Substituting in (1.16) we get

$$S^*_{(\ell)} = \sum_{K=1}^{N} p_K^{*2^\ell} = (\Sigma p_t^{2^\ell} - p_N^{2^\ell} + p_{N-1}^{2^\ell})\{1 + (p_N-p_{N-1}) + (p_N-p_{N-1})^2 + \ldots\}^{2^\ell} \qquad (3.2)$$

From (3.1) and (3.2) we get that p_i^* is of $O(N^{-1})$ and $S^*_{(\ell)}$ is of $O(N^{-2^\ell+1})$.

Hence it follows from (1.15) that W_K^* is of $O(N^{-K})$. As δ is of $O(N^{-1})$, it is evident from (1.13) that in order to evaluate π_{ij} to $O(N^{-4})$ we should first evaluate ϕ_{ij} to $O(N^{-4})$. Substituting from (3.1) and (3.2) into (1.14) we get correct to $O(N^{-4})$,

$$\phi_{ij} = 2p_ip_j[1+\{2(p_N-p_{N-1}) + \frac{p_ip_j}{\Sigma p_t^2}\} + \{\frac{2(p_N-p_{N-1})p_ip_j}{\Sigma p_t^2} + 3(p_N-p_{N-1})^2 + \frac{(p_N^2-p_{N-1}^2)p_ip_j}{(\Sigma p_t^2)^2} + \frac{p_i^3p_j^3}{\Sigma p_t^2 \Sigma p_t^4}\}]$$

for $1 \leq i \neq j > N$

and

$$\phi_{Nj} = 2p_{N-1}p_j[1+\{2(p_N-p_{N-1}) + \frac{p_{N-1}p_j}{\Sigma p_t^2}\} + \{\frac{2(p_N-p_{N-1})p_{N-1}p_j}{\Sigma p_t^2} + 3(p_N-p_{N-1})^2 + \frac{(p_N^2-p_{N-1}^2)p_{N-1}p_j}{(\Sigma p_t^2)^2} + \frac{p_{N-1}^3p_j^3}{\Sigma p_t^2\Sigma p_t^4}\}]. \qquad (3.3)$$

Substituting from (1.10) and (3.3) into (1.13) we get correct to $O(N^{-4})$,

$$\pi_{ij} = 2p_ip_j[1 + \frac{p_ip_j}{\Sigma p_t^2} + \{\frac{p_i^3p_j^3}{\Sigma p_t^2\Sigma p_t^4} + \frac{(p_N^2-p_{N-1}^2)p_ip_j}{(\Sigma p_t^2)^2} - (p_N^2 - p_{N-1}^2)\}] \text{ for } i \leq i \neq j < N$$

and

$$\pi_{Nj} = 2p_j(p_N-p_{N-1})\{1+(p_N+p_{N-1})+(p_N+p_{N-1})^2\} + 2p_{N-1}p_j[1 + \frac{p_{N-1}p_j}{\Sigma p_t^2} + \{\frac{p_{N-1}^3p_j^3}{\Sigma p_t^2\Sigma p_t^4} + \frac{(p_N^2-p_{N-1}^2)p_{N-1}p_j}{(\Sigma p_t^2)^2} - (p_N^2-p_{N-1}^2)\}] . \qquad (3.4)$$

Substituting from (3.4) into (1.2) we get correct to $O(N^0)$,

$$V(\hat{Y}_{H.T.})_B = \frac{1}{2}\Sigma p_t z_t^2 - \frac{1}{2}[\Sigma p_t^2 z_t^2 - \frac{(\Sigma p_t^2 z_t)^2}{\Sigma p_t^2}] - \frac{1}{2\Sigma p_t^2}[\Sigma p_t^4 z_t^2 - \frac{(\Sigma p_t^4 z_t)^2}{\Sigma p_t^4}] + \frac{1}{2\Sigma p_t^2}(p_N^2-p_{N-1}^2)(\Sigma p_t^2 z_t)[\frac{\Sigma p_t^2 z_t}{\Sigma p_t^2} - 2z_N] \qquad (3.5)$$

From (2.2) and (3.5) we get that the H.T. estimator has the same variance correct to $O(N^1)$ for either of the schemes A and B and is given by

$$V(\hat{Y}_{H.T.})_A = V(\hat{Y}_{H.T.})_B = \frac{1}{2}\Sigma p_t z_t^2 - \frac{1}{2}[\Sigma p_t^2 z_t^2 - \frac{(\Sigma p_t^2 z_t)^2}{\Sigma p_t^2}] \qquad (3.6)$$

Hartley and Rao have obtained the variance of the H.T. estimator for the procedure of Goodman and Kish and the variance correct to $O(N^1)$ is given by

$$V(\hat{Y}_{H.T.})_{G.K.} = \frac{1}{2}\Sigma p_t z_t^2 - \frac{1}{2}\Sigma p_t^2 z_t^2 \qquad (3.7)$$

Later it has been shown by Rao (1963, 1965) that the variance is given by (3.7) also for the procedures of Durbin (1953, 1967) and Yates and Grundy (1953). From this one might be tempted to conjecture that (3.7) infact holds for any πps procedure. (3.6) shows that this may not always be the case.

From (3.6) and (3.7) we have

$$V(\hat{Y}_{H.T.})_A - V(\hat{Y}_{H.T.})_{G.K.} = V(\hat{Y}_{H.T.})_B - V(\hat{Y}_{H.T.})_{G.K.} = \frac{1}{2}\frac{(\Sigma p_t^2 z_t)^2}{\Sigma p_t^2} \geq 0 \qquad (3.8)$$

From (3.6), (3.7) and (3.8) we conclude that

Theorem 1: When the variance of the corresponding H.T. estimator is considered to $O(N^1)$, Goodman and Kish procedure has a uniformly smaller variance than the Hanurav's procedure.

For N not sufficiently large the approximation to $O(N^1)$ of the variance may not be quite satisfactory and as such one might have to consider the approximation to $O(N^0)$. Variance of the H.T. estimator for the Goodman and Kish procedure derived by Hartley and Rao (1962) may be written as

$$V(\hat{Y}_{H.T.})_{G.K.} = \frac{1}{2}[\Sigma p_t z_t^2 - \Sigma p_t^2 z_t^2] - \frac{1}{2}[2\Sigma p_t^3 z_t^2 - \Sigma p_t^2 \cdot \Sigma p_t^2 z_t^2 - 2(\Sigma p_t^2 z_t)^2] \qquad (3.9)$$

Comparisons of (3.9) with either (2.2) or (3.5) is rather hard and may not lead to any positive conclusion. So in the next section we will try to compare both the procedures under a well known super population model.

4. Comparison of the two procedures under a super population model: In order to study the relative performance of different I.P.P.S. (Inclusion Probability Proportional to Size) schemes as measured by the variance of the corresponding H.T. estimator, it is convenient to assume some knowledge regarding the relationship between the variate y and the auxiliary characteristic x. Since unequal probability sampling is resorted to in the situations where y is approximately proportional to x it is reasonable to assume the model

$$y_i = \alpha + \beta x_i + e_i \qquad (4.1)$$

where α and β are unknown constants and e_i is a random variable such that $E(e_i|x_i) = 0$, $E(e_i^2|x_i) = ax_i^g$, $a \geq 0$, $g \geq 0$, and $E(e_i e_j|x_i, x_j) = 0$.

Theorem 2: Average variance of the corresponding H.T. estimator for any I.P.P.S. scheme under model (4.1) is

$$V^*(\hat{Y}_{H.T.}) = \alpha^2[\sum_i \frac{1}{\pi_i} + \sum_i \sum_{j(\neq i)} \frac{\pi_{ij}}{\pi_i \pi_j} - N^2] + ax^g(\frac{\Sigma p_t^{g-1}}{n} - \Sigma p_t^g) \qquad (4.2)$$

Proof: Taking the expectation of $V(\hat{Y}_{H.T.})$ under model 4.1 we get

$$V^*(\hat{Y}_{H.T.}) = \alpha^2[\sum_i \frac{1}{\pi_i} + \sum_i \sum_{j(\neq i)} \frac{\pi_{ij}}{\pi_i \pi_j} - N^2] + \alpha\beta[2\sum_i \frac{x_i}{\pi_i} + \sum_i \sum_{j(\neq i)} \frac{\pi_{ij}}{\pi_i \pi_j}(x_i + x_j) - 2Nx]$$

$$+ \beta^2[\sum_i \frac{x_i^2}{\pi_i} + \sum_i \sum_{j(\neq i)} \frac{\pi_{ij}}{\pi_i \pi_j} x_i x_j - x^2] + a[\sum_i \frac{x_i^g}{\pi_i} - \sum_i x_i^g],$$

which, upon using the relations $\pi_i = np_i$ and $\sum_{j(\neq i)} \pi_{ij} = (n-1)\pi_i$, reduces to (4.2).

Thus from (4.2) it follows that when $\alpha = 0$, the average variance of the corresponding H.T. estimator will be the same for all the I.P.P.S. schemes. However, if $\alpha \neq 0$, it can be observed from (4.2) that among all the I.P.P.S. schemes, the H.T. estimator corresponding to the scheme for which the value $\sum_i \sum_{j(\neq i)} \frac{\pi_{ij}}{\pi_i \pi_j}$ is least will have the least average variance. Thus a reasonable investigation will be to rank the various I.P.P.S. schemes according to the value of $\sum_i \sum_{j(\neq i)} \frac{\pi_{ij}}{\pi_i \pi_j}$ (= C, say). For this investigation we will confine to the case n = 2 only.

For the schemes of Durbin (1967), Yates and Grundy (1953), Durbin (1953), Goodman and Kish (1950) and for scheme A of Hanurav (1967) the approximate expressions for π_{ij} correct to $O(N^{-4})$ are respectively given by

$$\pi_{ij}^{(1)} = 2p_i p_j[1 + \{(p_i + p_j) - \Sigma p_t^2\} + \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3 - (p_i + p_j\Sigma p_t^2 + (\Sigma p_t^2)^2\}] \qquad (4.3)$$

$$\pi_{ij}^{(2)} = 2p_i p_j[1 + \{(p_i + p_j) - \Sigma p_t^2\} + \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3 + \frac{3}{4}p_i p_j - \frac{7}{4}(p_i + p_j)\Sigma p_t^2 + \frac{7}{4}(\Sigma p_t^2)^2\}] \qquad (4.4)$$

$$\pi_{ij}^{(3)} = 2p_i p_j[1 + \{(p_i + p_j) - \Sigma p_t^2\} + \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3 + p_i p_j - 2(p_i + p_j)\Sigma p_t^2 + 2(\Sigma p_t^2)^2\}] \qquad (4.5)$$

$$\pi_{ij}^{(4)} = 2p_i p_j[1 + \{(p_i + p_j - \Sigma p_t^2\} + \{2(p_i^2 + p_j^2) - 2\Sigma p_t^3 + 2p_i p_j - 3(p_i + p_j)\Sigma p_t^2 + 3(\Sigma p_t^2)^2\}] \qquad (4.6)$$

and

$$\pi_{ij}^{(5)} = 2p_i p_j[1 + \frac{p_i p_j}{\Sigma p_t^2} + \frac{p_i^3 p_j^3}{\Sigma p_t^2 \Sigma p_t^4}] \qquad (4.7)$$

Expressions (4.3) and (4.6) are from Asok and Sukhatme (1974) and expressions (4.4) and (4.5) are from Rao (1963). Using equations (4.3) to (4.7) and the relation $\pi_i = 2p_i$, the values of $\sum_i \sum_j \frac{\pi_{ij}}{\pi_i \pi_j}$ correct to $O(N^0)$ for the five schemes are respectively given by

$$C_1 = \frac{1}{2}[N(1 - N\Sigma p_t^2) + \{3N\Sigma p_t^2 + N^2(\Sigma p_t^2)^2 - 2N^2\Sigma p_t^3 - 2\}] \qquad (4.8)$$

$$C_2 = \frac{1}{2}[N^2 + N(1 - N\Sigma p_t^2) + \{\frac{3}{2}N\Sigma p_t^2 + \frac{7}{4}N^2(\Sigma p_t^2)^2 - 2N^2\Sigma p_t^3 - \frac{5}{4}\}] \qquad (4.9)$$

$$C_3 = \frac{1}{2}[N^2 + N(1 - N\Sigma p_t^2) + \{N\Sigma p_t^2 + 2N^2(\Sigma p_t^2)^2 - 2N^2\Sigma p_t^3 - 1\}] \qquad (4.10)$$

$$C_4 = \frac{1}{2}[N^2 + N(1 - N\Sigma p_t^2) - \{N\Sigma p_t^2 - 3N\Sigma(\Sigma p_t^2)^2 + 2N^2\Sigma p_t^3\}] \qquad (4.11)$$

and

$$C_5 = \frac{1}{2}[N^2 + \frac{1}{\Sigma p_t^2}(1-N\Sigma p_t^2)+\{\frac{(\Sigma p_t^3)^2}{\Sigma p_t^2 \cdot \Sigma p_t^4} - 1\}] \qquad (4.12)$$

It can be easily verified from (4.8) thru (4.11) that

$$C_1 \leq C_2 \leq C_3 \leq C_4 \qquad (4.13)$$

which is also a direct consequence of the comparisons made by Rao (1963, 1965) of the above four schemes without any model assumptions. From (4.11) and (4.12) we get

$$C_5 - C_4 = \frac{1}{2}[\frac{1}{\Sigma p_t^2}(1-N\Sigma p_t^2)^2+\{\frac{(\Sigma p_t^3)^2}{\Sigma p_t^2 \cdot \Sigma p_t^4} - 3N^2(\Sigma p_t^2)^2 + N\Sigma p_t^2 + 2N^2\Sigma p_t^3 - 1\}] \qquad (4.14)$$

Now, assuming $p_1, p_2, \ldots, p_N$ to be having a specific distribution Δ with moments μ_r' we can replace Σp_t^r in (4.14) by $N\mu_r'$ because we have from Khintchine's law of large numbers

$$\underset{N\to\infty}{\text{plim}}\ m_r' = \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\Sigma p_t^r = \mu_r' \ . \qquad (4.15)$$

In view of the relation $\Sigma p_t = 1$, we however should have

$$\mu_1' = \frac{1}{N} \qquad (4.16)$$

In the following we will investigate the relative efficiency of Hanurav's scheme A in relation to the other procedures mentioned here under various distributions of p_t.

Case (i) - χ^2 distribution: When the p_ts are distributed as $\frac{1}{\nu N}\chi^2_{(\nu)}$ where $\chi^2_{(\nu)}$ is the chi-square variate with ν degrees of freedom, from the relation

$$\Sigma p_t^r = N\mu_r' \qquad (4.17)$$

we get

$$\Sigma p_t^2 = \frac{\nu+2}{\nu N} \qquad (4.18)$$

$$\Sigma p_t^3 = \frac{(\nu+2)(\nu+4)}{\nu^2 N^2} \qquad (4.19)$$

and

$$\Sigma p_t^4 = \frac{(\nu+2)(\nu+4)(\nu+6)}{\nu^3 N^3} \ . \qquad (4.20)$$

Substituting these values in (4.14) we get

$$C_5 - C_4 = \frac{1}{2}[\frac{\nu N}{\nu+2}\{1 - \frac{\nu+2}{\nu}\}^2+\{\frac{\nu+4}{\nu+6} - \frac{3(\nu+2)^2}{\nu^2} + \frac{\nu+2}{\nu} + \frac{2(\nu+2)(\nu+4)}{\nu^2} - 1\}]$$

which after simplification reduces to

$$C_5 - C_4 = \frac{2N}{\nu(\nu+2)} + \frac{4(2\nu+3)}{\nu^2(\nu+6)} \geq 0 \qquad (4.21)$$

Case (ii) - β distribution: When the p_t's follow a beta distribution of the first kind with parameters $(\alpha_1 - 1, \alpha_2)$ where α_1 and α_2 are related by the equation

$$\mu_1' = \frac{\alpha_1}{\alpha_1+\alpha_2+1} = \frac{1}{N}$$

or

$$\alpha_2 = (N-1)\alpha_1 - 1 \qquad (4.22)$$

we get after substituting $N\mu_r'$ for Σp_t^r,

$$\Sigma p_t^2 = \frac{\alpha_1+1}{N\alpha_1+1} \qquad (4.23)$$

$$\Sigma p_t^3 = \frac{(\alpha_1+1)(\alpha_1+2)}{(N\alpha_1+1)(N\alpha_1+2)} \qquad (4.24)$$

and

$$\Sigma p_t^4 = \frac{(\alpha_1+1)(\alpha_1+2)(\alpha_1+3)}{(N\alpha_1+1)(N\alpha_1+2)(N\alpha_1+3)} \ . \qquad (4.25)$$

Substituting from (4.23) thru (4.25) in (4.14) we get

$$C_5 - C_4 = \frac{1}{2(\alpha_1+1)(\alpha_1+3)(N\alpha_1+1)^2(N\alpha_1+2)}[N\alpha_1^3(N^3 + 2N^2 - 7N + 4) + \alpha_1^2(3N^4 + 4N^3 - 22N^2 + 14N + 1) + \alpha_1(12N^3 - 33N^2 + 18N + 3) - 6(N-1)] \qquad (4.26)$$

Now,

$$N^3+2N^2-7N+4=N(N^2-1)+2N(N-3)+4 > 0 \text{ for } N \geq 3 \qquad (4.27)$$

$$3N^4+4N^3-22N^2+14N+1=N^3(3N-7)+11N^2(N-2)+14N+1 > 0 \quad \text{for } N \geq 3 \qquad (4.28)$$

and

$$\alpha_1(12N^3 - 33N^2 + 18N + 3) - 6(N-1) = (\alpha_1 - 1)[3N^2(4N-11) + 3(6N+1)] + 3N^2(4N-11) + 3(4N+3) > 0 \quad \text{for } N \geq 3, \qquad (4.29)$$

because $\alpha_1 > 1$.
(4.27) thru (4.29) implies that

$$C_5 - C_4 > 0. \qquad (4.30)$$

Case (iii) - Uniform distribution: When the p_t's follow a uniform distribution over the interval $(0, \frac{2}{N})$, we get from $\Sigma p_t^r = N\mu_r'$,

$$\Sigma p_t^2 = \frac{4}{3N},\ \Sigma p_t^3 = \frac{2}{N^2} \text{ and } \Sigma p_t^4 = \frac{16}{5N^3} \ .$$

Substitution of these values in (4.14) gives,

$$C_5 - C_4 = \frac{(4N - 3)}{96} > 0 \qquad (4.31)$$

In view of equations (4.13), (4.21), (4.30) and (4.31) it follows that when the variance is considered to $O(N^0)$, Hanurav's strategy would be inferior to those of Durbin (1967), Yates and Grundy (1953), Durbin (1953), and Goodman and Kish (1950) when the p_t's follow chi-square, beta or uniform distributions.

5. Numerical Illustration: The data presented in Table 1 is for the 20 districts in the Andhra Pradesh State of India. The x variable gives the populations (rounded off in thousands) in these districts as per the 1951 census and the y variable gives the exact population as per the 1961 census

Table 1

Population Figures for the 20 districts in the Andhra Pradesh State of India

i	X_i	Y_i
1	2123	2342291
2	2072	2288976
3	2301	2609311
4	1697	1978434
5	1736	2076103
6	2560	3009997
7	1794	2033963
8	1666	1913169
9	1628	1342140
10	1483	1764223
11	1617	1909644
12	1447	1590689
13	1821	2063601
14	1109	1226465
15	835	1021503
16	831	1009301
17	1428	1620417
18	1329	1545750
19	808	1057225
20	1287	1574797

In Table 2 are presented the variances correct to $O(N^2)$, $O(N^1)$ and $O(N^0)$ for the procedure of Goodman and Kish as well as Hanurav when samples of size 2 are considered. The variance correct to $O(N^2)$ for either of the procedures represents the true variance for the customary estimator in the case of probability proportional to size sampling with replacement. Values of the successive approximations indicate that the convergence is quite satisfactory inspite of the fact that the population size is much smaller than one usually encounters in practice. The relative difference between the two variances for larger sample sizes is however expected to be much higher than it is in this case.

Table 2

Approximations to $V(\hat{Y}_{H.T.})$

Order of approximation	Goodman and Kish procedure	Hanurav's procedure
$O(N^2)$	364525×10^7	364525×10^7
$O(N^1)$	347021×10^7	347068×10^7
$O(N^0)$	346217×10^7	346249×10^7

Acknowledgement: This work was supported in part under OEC Contract No. 0-73-6640 by the U.S. Office of Education, Department of Health, Education and Welfare.

References:

1. Asok, C. and Sukhatme, B. V. (1974). On Sampford's procedure of unequal probability sampling without replacement. Proceedings of the Social Statistics Section. American Statistical Association: 220-225.

2. Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. Journal of the Royal Statistical Society, Series B, 15: 262-269.

3. Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. Applied Statistics 16: 152-164.

4. Goodman, R. and Kish, L. (1950). Controlled selection - A technique in probability sampling. Journal of the American Statistical Association 45: 350-372.

5. Hanurav, T. V. (1966). An optimum πPS sampling. Annals of Mathematical Statistics 37: 1859.

6. Hanurav, T. V. (1967). Optimum utilization of auxiliary information: πPS sampling of two units from a stratum. Journal of the Royal Statistical Society, Series B, 29: 374-391.

7. Hartley, H. O. and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement. Annals of Mathematical Statistics 33: 350-374.

8. Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47: 663-685.

9. Rao, J. N. K. (1963). On three procedures of unequal probability sampling without replacement. Journal of the American Statistical Association 58: 202-215.

10. Rao, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. Journal of the Indian Statistical Association 3: 173-180.

11. Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. Biometrika 54:499-513.

12. Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. Journal of the Royal Statistical Society, Series B, 15:253-261.

POSTER SESSION ON FURTHER DEVELOPMENTS ON GRAPHICAL METHODS FOR STATISTICAL DATA DISPLAY

By Roberto Bachi (Hebrew University of Jerusalem, Israel)
and Vincent P. Barabba (Director, U. S. Bureau of the Census)

The poster session was organized in order to supplement the session on "graphical methods for presenting statistical data".

In the poster session examples were given of actual possibilities of progress in the graphical statistical field through avenues such as:

(a) clarification of some basic concepts of graphical presentation;

(b) reduction of limitations and shortcomings of classical methods;

(c) introduction of new graphical symbols.

The explanation was based on the following exhibits (prepared at the Hebrew University of Jerusalem, apart from 2.1-2.2 and 4.4).

1. New Univariate Symbols

1.1 Examples of Graphical Rational Patterns (GRP)

1.2 Comparison of GRP method and diagrammatic method

1.3 - 1.4 GRP representing double-entry tables

1.5 A function of X and Y shown by GRP

1.6 Cumulated percentages (Y) of Women of Age X, with number of children Z (shown by GRP)

1.7 Circles proportional to population of towns over a map showing rural densities by GRP

1.8 GRP showing traffic density

1.9 Scale of compact GRP 1 - 99

1.10 Compact GRP proportional to population of SMSA

1.11 - 1.12 Compact GRP showing percentages of Blacks over the population of SMSA. In 1.12 the area of the basic circle is proportional to total population, and that of the compact GRP is proportional to Black population.

2. New Bivariate and Multivariate Colored Symbols

2.1 - 2.2 Examples of the method of blending colors for the presentation bivariate maps (Bureau of the Census)

2.3 - 2.5 Bivariate colored symbol for SMSA located over a map of the USA

2.6 - 2.7 Bivariate colored symbols for SMSA over a scatter diagram showing two additional variables

2.8 Multivariate colored symbols for SMSA over a scatter diagram showing two additional variables

3. Improving the Common Diagram

3.1 Linking the scales of X and Y in diagrams

3.2 Modified frequency polygon compared to a histogram and to a common frequency polygon.

3.3 Diagram showing magnitudes and absolute and relative variations

4. Graphing Regional Averages

4.1 Map of median income by states of the USA, obtained by covering the area of each state by one out of a few patterns

4.2 Map 4.1 re-drawn by covering the area of each state by a repeated GRP

4.3 Map 4.2 re-drawn by giving in each state a number of GRP proportional to the population

4.4 Percentage of area urbanized in each county of California shown by Tobler's scale (Bureau of the Census)

As most of the exhibits were in color, they cannot be reproduced here. A verbal description of graphs would also be impractical.

Therefore, the reader is referred to a detailed publication which is being prepared at present by the Bureau of the Census. In this publication all the exhibits will be reproduced, together with the detailed explanations and papers given at both sessions on graphical methods.

AN INTERVIEWER VARIANCE STUDY FOR THE EIGHT IMPACT CITIES OF THE NATIONAL CRIME SURVEY CITIES SAMPLE

Leroy Bailey, Thomas F. Moore, and Barbara A. Bailar, U.S. Bureau of the Census

1. Introduction and Background

In July 1972, the Bureau of the Census began conducting a survey for the Law Enforcement Assistance Administration in eight central cities (National Crime Survey - Central Cities Sample). The survey was designed to gather data relating to personal crime. The sample consisted of approximately 12,000 housing units from each of the cities designated by the U.S. Department of Justice as "impact aid cities"--Atlanta, Baltimore, Cleveland, Dallas, Denver, Newark, Portland, and St. Louis. These units were surveyed again during the first half of 1975; it is the data relating to this period on which the Interviewer Variance Study (IVS) is based.

The principal objective of the National Crime Survey (NCS) is to obtain estimates of the extent of victimization attributable to the major crimes of assault (including rape), burglary, larceny, auto theft, and robbery.

2. Description of Study

It is generally accepted that the major part of response variance in small areas is the interviewer's contribution. In the NCS, one interviewer usually works in a specified assignment area of about 80 housing units. It has been a concern that there may be large differences between interviewers in their interviewing methods or in the application of their training instructions, which could possibly affect the data. The major intent of the IVS was to obtain estimates of the contribution of interviewers to the correlated response variance of NCS statistics, specifically the victimization rates for some of the major crimes on which the survey focuses.

In order to estimate the correlated component of response variance attributable to interviewers, the method of interpenetrated subsamples originally described by Mahalanobis [5] was used. In each of the eight impact cities, 144 interviewer assignment areas and 18 crew leader districts, each containing eight geographically contiguous interviewer assignment areas were delineated. Then, within each crew leader district, pairs of interviewer assignment areas were formed.

Interviewers were assigned to crew leader districts and interviewer assignment areas based on the geographical proximity of their homes to each of the areas. A random selection of 36 assignment pairs from each impact city was then made in Washington. Finally, within each pair of assignment areas, housing units were assigned in a systematic method, so that each interviewer was assigned approximately half of the units. For each city, this procedure permitted the comparison of the work of pairs of interviewers and the estimation of an average correlated response variance for an NCS interviewer assignment area.

3. The Mathematical Model

The mathematical model used in this study is that developed by Hansen, Hurwitz, and Bershad [4] and used in several other studies previously conducted by the Bureau. The estimator of the sampling variance, though it contains the simple response variance, does not reflect the correlated response variance. Since correlated response variance includes the effect of interviewer differences in the understanding and application of the training instructions in the conduct of the survey, this component could be an important source of variability.

An estimator of the total variance, T^2, is as follows:

$$T^2 = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^2 \qquad (1)$$

where $\bar{x}_1$ is the mean based on the work of interviewer 1 and $\bar{x}_2$ is the mean based on the work of interviewer 2. This includes not only the sampling variance, but the simple response variance, the correlated response variance, and the covariance between response and sampling deviations. We needed an estimator of the sampling variance and used the following:

$$S^2 = \frac{\sum_{h}^{2}\sum_{j}^{n}(x_{hj} - \bar{x}_h)^2}{2n(n-1)} \qquad (2)$$

where x_{hj} is the value of the j-th unit assigned to the h-th interviewer and $\bar{x}_h$ is the mean for the h-th interviewer.

S^2 is the average variance for a mean based on a simple random sample of n units, and averaged over the two interviewers. The expected value of S^2 shows that it also includes the simple response variance.

Subtracting S^2 from T^2, provides an estimator of the correlated response variance. Thus, for each pair of assignment areas, there was available an estimate of the total variance, the sampling variance, and the correlated response variance for each statistic. We then averaged these values over all pairs of assignment areas within the impact city.

Most of the statistics from the crime survey are not simple means, but are rates or ratios of means. For a ratio, $r = \bar{x}/\bar{y}$, the total relative variance was computed as follows:

$$T_r^2 = \frac{a\sum_i^a T^2_{\bar{x}_i}}{(\sum_i^a \bar{x}_i)^2} + \frac{a\sum_i^a T^2_{\bar{y}_i}}{(\sum_i^a \bar{y}_i)^2} - \frac{2a\sum_i^a T_{\bar{x}\bar{y}_i}}{\sum_i^a \bar{x}_i\bar{y}_i} \qquad (3)$$

where a is the number of assignment areas, and for the i-th area,

$T^2_{\bar{x}_i}$ is the total variance for the mean for characteristic X;

$T^2_{\bar{y}_i}$ is the total variance for the mean for characteristic Y;

$T_{\bar{x}\bar{y}_i}$ is the total covariance between the means of characteristics X and Y;

$\bar{x}_i$ is the mean for characteristic X; and

$\bar{y}_i$ is the mean for characteristic Y.

Similarly, the sampling relative variance was computed and the difference between them was used as an estimate of the correlated response relative variance. These are estimates that are shown in column (5) of Table 1.

Because these estimates of variability are based on a sample, they are themselves subject to sampling error. We estimated the variability of these estimates by considering each assignment area as an ultimate cluster. Suppose we wanted to estimate the variability of T_r^2 as shown in equation (3). We estimated T_r^2 for each assignment area and denoted these values by t_i. Then the estimate of variance was:

$$s^2 = \frac{\sum_i^a (t_i - T_r^2)^2}{a(a-1)} \quad . \qquad (4)$$

The square roots of the averages of these values over the eight impact cities are shown in the last column of Tables 2a-2d.

4. Summary of Results

Included among the major statistics computed for the IVS were estimates of the total variance, sampling variance, and correlated response variance for numerous NCS items. The corresponding estimates of relvariance were also computed. An illustration of the manner in which these estimates have been displayed is provided by table 1, which shows data from Baltimore. It should be noted that estimates contained in this table and in those which follow were derived from weighted data. The base (4,823 persons) which appears directly above the table is a weighted estimate of the average number of persons per interviewer assignment area in Baltimore. The corresponding unweighted base is 175 persons. Columns (3), (4), and (5) are respectively the total relvariance, sampling relvariance, and correlated response relvariance for the major personal victimization rates. For these rates, column (6) gives the ratios of correlated response variance to sampling variance. Such ratios represent the relative increase in variability due to the contributions of interviewers. A more detailed discussion of the IVS data from Baltimore will be presented in conjuction with data from the other impact cities.

In order to avoid a paper of unreasonable length, the only estimates which have been provided which pertain to all of the eight impact cities are ratios of correlated response variance to sampling variance for a specific set of personal and household victimization rates.

In table 2a estimated ratios of correlated response variance to sampling variance are given for the major personal victimization rates. Exclusive of those for Cleveland and Denver, the ratios for the overall or total personal victimization rate are equal to or exceed .50. The 1.40 ratio for Newark is exceptionally large. A ratio of this magnitude means that the estimated variance of this item is considerably understated. To get the total variance, one must multiply the sampling variance by one plus this ratio. For example, in Newark, the sampling variance of the total rate must be multiplied by 2.40 to reflect the total variance of this statistic.

The disparity within the cities between the ratios for the major subcategories is apparent. The ratios for assaultive violence are generally higher than those for personal theft without assault. Within the cities, differences can also be detected between the major groups of which the assaultive violence item is comprised. Interviewers appear to have a considerably greater influence on the variability of the victimization rates for "assaultive violence without theft" than for "assaultive violence with theft."

The generally large ratios for the overall victimization rate, as well as for assaultive violence and assaultive violence without theft, suggest that there may have been differences in the method in which NCS interviewers applied the concepts or definitions directly related to a determination of the incidence of assaultive violence without theft. Perhaps the variability among the interviewers may reflect the manner in which they view assaultive violence involving acquaintances or relatives and friends.

In table 2b ratios of correlated response variance to sampling variance are given by the race of the respondent for the major personal victimization rates. The ratios for the "other" category are generally very small and have large sampling errors. Although their sampling errors are also sizable, the other estimates provided in the table are more reliable.

The contention that large interviewer variability is primarily associated with respondents belonging to a specific racial group is not strongly supported by the entries of table 2b. In three of the cities the estimated ratios for the overall victimization rate are higher for whites than for blacks, while in the other five the ratios are higher for blacks. A similar pattern exists for the other four categories. In addition, ratios similar to those which are evident in table 2a are exhibited for blacks in Atlanta, Dallas, and Newark; however in Denver and St. Louis the ratios for the whites compare favorably with those for the combined rates in table 2a.

Since the overall ratio for whites is larger than that for blacks in Portland and Denver and with the exception of these two cities all of the impact cities have a black population which is at least 25 percent of the total population, it might be suspected that in the cities with the larger black populations, the ratios for blacks would be larger. However, in St. Louis, a city whose black population exceeds 40 percent, the overall ratio is larger for whites, while in Dallas, where about 25 percent of the population is black, the ratio is larger for blacks.

The ratios presented in table 2c relate to the major household victimization rates. They indicate that the household victimization rates for a number of items, like the personal rates, are affected significantly by interviewers. The ratios for the total household rate are sizable for all of the cities, with four of them being in excess of .50 and one greater than 1.00. Similarly, the ratios for "larceny under $50" were also fairly large. In contrast, the estimates for the burglary and auto theft items were usually quite small.

As was expected, these ratios vary among the cities. For Denver and Cleveland they tend to be lower than those for the other cities, as was the case with the ratios involving the personal victimization rates; the ratios for Newark and Atlanta are again among the highest reported for the eight cities.

Table 2d, which is the last of the tables relating to all eight cities, provides ratios by race of household head for the overall household victimization rates.

It is interesting to observe that in the five cities where the black population is about 40 percent or more of the total population, the ratios are larger for this group. In addition, in the cities where blacks are in the majority (Newark and Atlanta) the combined (both sexes) ratios are largest. Perhaps this finding may be related to differences in the races of the interviewers and the respondents and to the possibility of their having varying perceptions of household crimes. Again, the number of sample cases for the "other" category is much too small to provide adequate estimates.

In spite of the large household ratios for blacks, the ratios for whites in all of the cities except Newark and St. Louis were sizable. This result is consistent with that involving personal victimizations, and it strengthens the argument against the suggestion that large interviewer effects are peculiar to households headed by a person of a particular race.

Far more data from the IVS are available than are included in the tables which accompany this paper. In addition to assessing the effects of interviewers on estimates of certain personal and household victimization rates, the IVS provided estimates of the component of response variance attributed to interviewers for labor force, occupation, income, and education attainment items. For many of these items large estimates of correlated response variance were also observed. Additional reports which will provide more comprehensive information relevant to the study will be forthcoming.

5. Limitations of Data

Several cautions must be emphasized regarding inferences which might be drawn from the IVS data. Initially, it must be remembered that the IVS provided estimates of correlated response variance which are only applicable to the eight impact cities and to areas about the size of the average interviewer assignment area for the respective cities. Therefore, these estimates should not be used definitively to assess the quality of crime statistics in other cities or for the total United States.

Secondly, measurement errors attributable to interviewers reflect general survey conditions which affect their individual performances. Included among such conditions are the qualifications and training of the interviewers, their salaries, the frequency with which data are collected from sample units, the survey's recall period, and the organization selected to collect the data. Consequently, it is inappropriate to compare IVS statistics to statistics from similar studies for which these "controllable" general conditions differ substantially from those of the IVS.

Finally, another problem which limits the IVS data, but probably not to a great extent, is the fact that the actual conduct of the study deviated somewhat from the experimental design. Despite three levels of control, some interviewers completed units originally assigned to another interviewer. These violations undermined the objectives of randomizing the interviewer assignments, and forced the deletion of some of the sample units.

6. Conclusion and Recommendations

Two general inferences can be immediately drawn from the data which have been presented.

1. The 1975 NCS statistics for the eight impact cities are subject to interviewer variability. The extent to which interviewers influenced these statistics varied considerably among the cities and according to the nature of the statistics.
2. Data users may develop misconceptions regarding the quality of NCS data. To the extent that response variances are sizable, the total variances of these estimates are understated. Caution should therefore be exercised regarding the interpretation of differences among the impact cities in the reported victimization rates. These differences may be obscured by sampling and nonsampling variances.

What can be done to adequately assess the effects of interviewers on NCS statistics and to eventually reduce such effects? Obviously, researchers should initially thoroughly review the data from the IVS and other related research to attempt to more accurately categorize the effects of interviewers on the data.

Secondly, the training and observation of NCS interviewers could possibly be revised so that greater emphasis is placed on the concepts related to items for which large interviewer variability is reported.

Thirdly, some form of additional research on the interviewers is recommended. Perhaps interviewers with certain characteristics or qualifications are more prone to commit errors. These interviewers may require supplemental training and more observation.

Finally, more studies similar to the IVS are suggested, so as to provide a number of estimates for the same statistic under different conditions. Hopefully some or all of these procedures would contribute to the development of "optimum" sample designs for the NCS-related surveys.

References

[1] Bailar, B.A., "Some Sources of Error and Their Effect on Census Statistics", presented at the Population Association Meetings in Seattle, Washington, April, 1975.

[2] Bailar, B.A., "Estimators for the Interviewer Variance Study: National Crime Survey", Unpublished Paper, Bureau of the Census, Washington, D.C., February, 1976.

[3] Bailey, L. "Toward A More Complete Analysis of the Total Mean Square Error of Census and Survey Statistics", Proceedings of The Social Statistics Section, ASA, 1975, pp. 1-10.

[4] Hansen, M.H., Hurwitz, W.H., and Bershad, M.A., "Measurement Errors in Censuses and Surveys", Bulletin of the International Statistical Institute, Vol. 38, No. 2, 1961, pp. 359-374.

[5] Mahalanobis, P.C. "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," Journal of the Royal Statistical Society, Vol. 109, Part 4, 1946, pp. 326-378.

[6] Thompson, M.M. "Recent Developments in the National Crime Panel," presented at a meeting of the Washington Statistical Society, March, 1976.

[7] U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses Population and Housing, 1960: Effects of Interviewers and Crew Leaders, Series ER60, No. 7, Washington, D.C., 1968.

TABLE 1.1/--ESTIMATED SAMPLING RELVARIANCES, CORRELATED RESPONSE RELVARIANCES, AND TOTAL RELVARIANCES FOR MAJOR PERSONAL VICTIMIZATION RATES FOR BALTIMORE
(Base for rates - 4,823 Persons2/)

Kinds of victimizations	Number 2/ of victim-izations (1)	Rates (Victim-izations per person) (2)	Relvariances			Ratios of correlated response to sampling variance (6)
			Total (3)	Sampling (4)	Correlated response (5)	
Total	525	.1088	.1586	.0989	.0598	.60
Assaultive violence total	292	.0605	.2625	.1612	.1013	.63
Assaultive violence with theft	55	.0115	.7512	.6452	.1060	.16
Assaultive violence with-out theft	236	.0490	.2943	.1915	.1028	.54
Personal theft without assault	233	.0484	.2211	.1711	.0500	.29

1/ These estimates are applicable to an area approximately equal in size to an average NCS interviewer assignment area in Baltimore.

2/ These numbers are weighted counts.

TABLE 2a.--RATIO OF CORRELATED RESPONSE VARIANCE TO SAMPLING VARIANCE FOR THE 1975 IMPACT CITIES (Personal Victimizations)

Kinds of personal victimizations	Atlanta (1)	Baltimore (2)	Cleveland (3)	Dallas (4)	Denver (5)	Newark (6)	Portland (7)	St. Louis (8)	Average estimated standard error (9)
Total	.69	.60	.18	.50	.34	1.40	.59	.98	.33
Assaultive violence	.70	.63	.31	.45	.09	.85	.68	1.17	.28
Assaultive violence with theft	.00	.16	.33	.00	.00	.10	.24	.00	.15
Assaultive violence without theft	.79	.54	.37	.50	.12	1.21	.70	1.17	.27
Personal theft without assault	.27	.29	.00	.22	.47	.79	.27	.07	.25

TABLE 2b.--RATIO OF CORRELATED RESPONSE VARIANCE TO SAMPLING VARIANCE FOR THE 1975 IMPACT CITIES (Personal Victimizations By Race)

Kinds of personal victimizations	Atlanta (1)	Baltimore (2)	Cleveland (3)	Dallas (4)	Denver (5)	Newark (6)	Portland (7)	St. Louis (8)	Average estimated standard error (9)
Total									
White	.39	.17	.00	.29	.40	.64	.54	1.20	.25
Black	.90	.42	.01	.71	.00	1.13	.00	.19	.23
Other	.00	.00	.00	.00	.02	.64	.04	.00	.17
Assaultive violence									
White	.33	.48	.09	.35	.08	.11	.61	1.37	.23
Black	.83	.08	.24	.41	.00	.63	.36	.38	.18
Other	.08	.00	.00	.00	.07	.49	.08	.00	.18
Assaultive violence with theft									
White	.00	.23	.00	.00	.00	.00	.10	.20	.12
Black	.21	.00	.58	.01	.00	.14	.00	.00	.14
Other	.08	.00	.00	.22	.00	.59	.02	.97	.18
Assaultive violence without theft									
White	.28	.34	.32	.47	.08	.56	.66	1.14	.24
Black	.97	.04	.00	.34	.09	1.05	.49	.63	.18
Other	.08	.05	.00	.00	.12	.13	.08	.00	.21
Personal theft without assault									
White	.63	.00	.00	.07	.43	.60	.32	.00	.21
Black	.00	.44	.00	.00	.13	.79	.00	.10	.17
Other	.00	.00	.00	.22	.00	.00	.00	.97	.17

TABLE 2c.--RATIO OF CORRELATED RESPONSE VARIANCE TO SAMPLING VARIANCE FOR THE 1975 IMPACT CITIES (Household Victimizations)

Kinds of household victimizations	Atlanta (1)	Baltimore (2)	Cleveland (3)	Dallas (4)	Denver (5)	Newark (6)	Portland (7)	St. Louis (8)	Average estimated standard error (9)
Total	1.21	.76	.40	.37	.43	1.00	.56	.40	.38
Burglary, forcible entry something taken	.01	.00	.09	.44	.10	.00	.00	.28	.22
Burglary, unlawful entry without force	.00	.70	.04	.56	.02	.08	.35	.00	.21
Burglary, attempted forcible entry	.64	.07	.00	.21	.00	.00	.17	.17	.20
Larceny under $50	1.00	.55	.31	.64	.37	.60	.64	.41	.34
Larceny $50 or more	.57	.32	.28	.29	.28	.10	.55	.56	.27
Larceny NA amount	.45	.00	.00	.48	.00	.39	.03	.00	.15
Attempted larceny	.41	.48	.36	.00	.14	.42	.31	.38	.20
Auto theft, theft of car	.00	.13	.13	.00	.00	.00	.04	.40	.18
Auto theft, attempted theft of car	.38	.18	.00	.07	.20	.00	.42	.00	.12

TABLE 2d.--RATIO OF CORRELATED RESPONSE VARIANCE TO SAMPLING VARIANCE FOR THE 1975 IMPACT CITIES (Household Victimizations by Race)

Race of household head	Atlanta (1)	Baltimore (2)	Cleveland (3)	Dallas (4)	Denver (5)	Newark (6)	Portland (7)	St. Louis (8)	Average estimated standard error (9)
White	.70	.96	.41	.68	.85	.32	.97	.14	.33
Black	2.20	1.28	.85	.62	.00	1.65	.58	.75	.39
Other	.00	.13	.03	.00	.35	.57	.37	.00	.13

EVALUATION OF THEORIZED FACTOR STRUCTURE OF THE MMPI FOR MALE AND FEMALE POPULATIONS*

Barbara M. Barker and Harry R. Barker
University of Alabama

Introduction

Many tests require different scoring keys for males and females. In doing so the test manufacturers are implying that, along the dimensions that the tests measure, males and females are not represented adequately by the same items or by items scored in the same direction. The women's movement, on the other hand, has been trying to break down sex role stereotypes and asserts that women and men share the same characteristics along any continuum.

This paper looks at four current theories about the factor structure of the Minnesota Multiphasic Personality Inventory (MMPI) as related to the issue of male-female differences. The four theories tested are the conventional scoring of the clinical scales (Hathaway & McKinley, 1951), Lushene's (1967) theory, Tryon's (1966) theory, and the factor scale theory of Barker, Fowler, and Peterson (1971).

The Lushene theory provides different subscales for males and females using 18 subtests for males and only 9 for females. The clinical scales score all subtests with the same key for males and females except the M-F scale. This scale postulates differences accounted for by sex. The Tryon theory scores subtests with one key for both males and females, making no distinction between the sexes. The Barker et al theory, because it originated from a male population, makes no provision for females at all.

Various methods of analysis were used in arriving at the theories for the MMPI examined here. The clinical scales were constructed through item analysis and subjective content validity. Tryon's theory involved factoring random subsets of items and converging on salient clusters. This theory was derived from 310 adult subjects, sex not specified. At no time were all 566 items in the same analysis. Lushene used an obverse factor analysis method to factor subjects (people) rather than items or variables. Lushene factored separately 189 males and 253 females. Because of the small number of subjects used, the subscales are of questionable validity. The Barker et al theory was based on a direct factor analysis of 373 items from a short-form of the MMPI. Subjects were 1575 male hospitalized VA patients. The 373 items contained all conventional clinical scale items. To date, the 566-item MMPI has been too large to factor directly, and thus, none of these theories is based upon the representation of all items in a conventional factor analysis.

Horst (1965) proposed an indirect factor analysis method which can be applied to very large numbers of variables. Horst's indirect method was implemented by Barker in 1971 in a computer program, CORR99. The program has been revised and tested over more than five years (Sloan, 1973; Stallings, 1973; Barker & Barker, 1975), and provides a method for testing theories about large data matrices (Barker & Barker, 1976a; Hamlett, 1976). This paper proposes to evaluate by Horst's indirect method the four theories of the MMPI scales as they relate to women, and then to compare the theories with the male counterparts. The best theory for females will be compared through rotation to the equivalent male theory.

Methodology

The original MMPI standardization sample was used for data. The 511-item MMPI was administered to 315 female friends and relatives of hospitalized neuropsychiatric patients. Fifty-five items were subsequently added to the scale and identified in the theories. Since these items were not present in the original testing, they were not considered in the analyses. The MMPI items were factor analyzed by the indirect factor-analytic method (CORR99, Barker, 1973) on a Univac 1110 computer. Items with a factor load equal to or greater than .30 (positive or negative) on only one factor were used.

Each of the four theories was tested separately and the information measure D (relative uncertainty reduction was computed. The D measure is used to relate the degree of association between theorized factor structure and actually obtained factor structure. The paper by Barker and Barker (1976b) supplied the computed D measures from each of the theories for the males. Differences in adequacy among the four theories for males and females were compared. The results of the theory supplying the most accurate

*The research was made possible by funds from the Research Grants Committee of the University of Alabama.

results for women were used as a reference criterion to which the equivalent male theory results were rotated (CORR22).

Results

Matrices for the four D measures are shown in Tables 1 through 4. For the D measure, 1.00 is a perfect certainty or prediction capability. To the extent that more uncertainty or error in prediction exists, the D measure will approach 0.00. Relative ranking of the D measures from least to most agreement for males and females are:

D measure for	Females	Males
1. Conventional	.32	.42
2. Tryon	.42	.42
3. Lushene	.43	.56
4. Barker et al	.59	.52

The low D measures for the females from these four theories would seem to indicate little agreement between existing theory and the actual factor structure for the 566 item instrument. Relatively, the theory for women based on the factored 373-item instrument was the most accurate. If one accepts the view that the factor structure should differ for male and female, this is puzzling in that there were no females in the Barker et al sample. The rank ordering for males differed from females with the Lushene theory being slightly more adequate than the factor-based theory of Barker et al. The results from Tryon's theory produced identical D measures for males and females which one could postulate as he does not differentiate between the sexes. The conventional clinical scales and Lushene's theory appear to define male dimensions better than female. Since the Barker et al theory was so much better than the Lushene theory for females, and because of the different numbers of factors in Lushene's theory for male and female, the results of the Barker et al solutions for males and females were rotated to maximum alignment to determine the similarity of the factor solution for the two sexes. Results of this rotation are given in Table 5.

Approximately two-thirds of the cosines between items for males and females were over .70 indicating much agreement between the two solutions. If the basic factor structure is the same for male and female, it is conceivable that the female factor solution is more accurate than that obtained for the males. This seems reasonable due to the larger and more adequate sampling of females (females = 315; males = 225).

An inspection of the cosines showed some negatively related items. This would indicate that males and females tended to answer in opposite directions on these items. There appeared to be no apparent patterning on subscales in these differences. A few representative items are:

46. My judgment is better than it ever was.
183. I am against giving money to beggars.
113. I believe in law enforcement.
376. Policemen are usually honest.
199. Children should be taught all the main facts of sex.
176. I do not have a great fear of snakes.
522. I have no fear of spiders.
454. I could be happy living all alone in a cabin in the woods or mountains.

Summary

Four current theories about the factor structure of the MMPI as they relate to male-female differences were compared by an indirect factor analytic method. None of the theories proved adequate in estimating the factor structure of the 566-item MMPI.

The most impressive theory for females was the Barker et al theory with a D measure of .59. For the males, the Lushene theory with a D measure of .56 was virtually identical in adequacy to the Barker et al theory (D = .52). Factor results for females, based on the Barker et al theory, were used as a reference criterion to which the male factors of the same theory were rotated.

References

Barker, H. R. Behavioral Sciences Statistics Program Library. University of Alabama: Reproduction Services, 1973.

Barker, H. R., and Barker, B. M. Summary of research on a novel, indirect factor analytic solution. Proceedings of the Social Statistics Section, American Statistical Association, 1975, 298-301.

Barker, H. R., and Barker, B. M. An indirect factor method for testing the dimensionality of large data sets. Paper presented at Southeastern Psychological Association, New Orleans, La., March 19, 1976.(a)

Barker, H. R., and Barker, B. M. An indirect method for testing the dimensionality of large data sets. Paper presented at American Statistical Association, Boston, Mass., August 23, 1976.(b)

Barker, H. R., Fowler, R. D., and Peterson, L. D. Factor analytic structure of the short form MMPI items in a VA hospital population. Journal of Clinical Psychology, 1971, Vol. XXVII, No. 2, 228-233.

Hamlett, C. C. Validation of the theorized factor structure of the Personal Orientation Inventory. Paper presented at American Statistical Association, Boston, Mass., August 23, 1976.

Hathaway, S. R., and McKinley, J. C. MMPI Manual, Revised Ed. New York: Psychological Corp., 1951.

Horst, P. Factor analysis of data matrices. New York: Holt, 1965.

Lushene, R. E. Factor structure of the MMPI item pool. Unpublished master's thesis, Florida State University, 1967.

Sloan, H. C. A simplified procedure for estimating the factor structure of large data matrices. Unpublished doctoral dissertation, University of Alabama, 1973.

Stallings, N. A. Evaluation of an indirect method of estimating the factor structure of large data matrices. Unpublished doctoral dissertation, University of Alabama, 1973.

Tryon, R. C. Unrestricted cluster and factor analysis, with applications to the MMPI and Holzinger-Harman problems. Multivariate Behavioral Research, 1966, Vol. 1, 229-244.

Table 5.

Frequency Distribution of Cosines

Cosines	Freq.	Cum. Freq.	Cum. %ile
.90+	92	92	18
.80+	152	244	48
.70+	88	332	65
.60+	66	398	78
.50+	43	441	86
.40+	23	464	91
.30+	18	482	94
.20+	8	490	96
.10+	8	496	97
.01+	13	511	100

TABLE 1

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (CONVENTIONAL THEORY)

ITEM SUBSETS (X)	Factors (Y) I	II	III	IV	V	VI	VII	VIII	IX	False -	Sum
1	8										8
2					2		3	2		16	23
3							7			9	16
4				2			2			17	21
5					4		1			16	21
6						3	2			12	17
7							14			4	18
8							4	4		30	38
9							2			17	19
False +	3				1		42	2			48
Sum	11			2	7	3	77	8		121	229

H(X) = 3.473   H(X,Y) = 4.043

H(Y) = 1.690   HT = 1.121   D = .323

TABLE 2

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (TRYON THEORY)

	Factors (Y)							False	
X ITEM SUBSETS	I	II	III	IV	V	VI	VII	-	Sum
1	23							2	25
2		19						13	32
3			2		4			19	25
4	1			4	1		2	19	27
5					12			8	20
6					5		1	16	22
7							23	12	35
False +		4		1	8		3		16
Sum	24	23	2	5	30		29	89	202

H(X) = 2.962 H(X,Y) = 3.981

H(Y) = 2.252 HT = 1.232 D = .416

TABLE 3

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (LUSHENE THEORY)

	Factors (Y)									False	
X ITEM SUBSETS	I	II	III	IV	V	VI	VII	VIII	IX	-	Sum
1	19	1	1		1					45	67
2		26								22	48
3			17							10	27
4				3						17	20
5					6					6	12
6						4				7	11
7							5			2	7
8								2		1	3
9	1								3		4
False +	17	12	5				2				36
Sum	37	39	23	3	7	4	7	2	3	110	235

H(X) = 2.817 H(X,Y) = 3.918
H(Y) = 2.312 HT = 1.211 D = .430

TABLE 4

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS
(BARKER, FOWLER, AND PETERSON THEORY)

ITEM SUBSETS X	Factors (Y) I	II	III	IV	V	VI	VII	VIII	IX	False -	Sum
1	31									11	42
2	1	25								23	49
3	2		9							16	27
4				9						13	22
5					13					1	14
6	4					6				8	18
7	1						7			25	33
8	1							10		1	12
9									3	5	8
False +	51	7	4			1	1	1			65
Sum	91	32	13	9	13	7	8	11	3	103	290

H(X) = 3.072   H(X,Y) = 3.748

H(Y) = 2.483   HT = 1.807   D = .588

AN INDIRECT METHOD FOR TESTING THE DIMENSIONALITY OF LARGE DATA SETS

Harry R. Barker and Barbara M. Barker, University of Alabama

The modern computer imposes severe limitations on the number of variables which can be factored by conventional methods. Various alternative factor solutions are discussed and evaluated by Barker, Fowler, and Peterson (1971). Each of these alternatives proves to be grossly inadequate from a theoretical standpoint.

An alternative factor solution proposed by Horst (1965) seems theoretically sound and offers a method of factoring huge data matrices with minimum time requirements on a medium size computer. Horst's method essentially entails combining the variables into a relatively small number of subsets, factoring the subsets and then estimating the factor loads of the individual variables. In this way, the large matrix of variables is bypassed. (See Barker and Barker (1974) for a more extended illustration of the method.)

Horst's indirect factor method has been systematically studied and refined over an extended period of time. See Stallings (1973), Sloan (1973), Barker and Barker (1974), Barker and Barker (1975-A), Barker and Barker (1975-B), Barker, Barker, and Carlton (1975), Barker and Barker (1975-C), Barker and Barker (1976). A summary review of research on the method through 1975 is given by Barker and Barker (1975-C). Two major findings resulted from the series of studies:

(1) In order for Horst's method to produce accurate results it is essential that the variables comprising each subset belong to the same factor.
(2) Combination of the variables into a subset, by totaling, must take into account the directionality of the measurement, positive or negative.

These findings appeared, at first, to impose drastic limitations on the use of Horst's method. For it is essential to know the correct solution before the method can produce the correct solution. On further reflection it was noted that theory serves just this purpose; providing an assertion as to what the correct answer or solution is. Horst's method then seems to be an ideal method for testing the accuracy of theory. The method has proved useful in evaluating the accuracy of competing theories as to the factor structure of large matrices. (See Barker and Barker (1976) and Hamlett (1976). The manner of using the procedure is as follows. A theory is used to cluster variables into subsets. Horst's factor method is then applied and one simply notes the extent to which variables which were previously clustered into the same subset emerge together on a single factor.

The purpose of this study is to demonstrate the use of Horst's indirect factor method in evaluating four competing theories relating to the factor structure of the widely used Minnesota Multiphasic Personality Inventory (MMPI).

The four competing theories to be evaluated are as follows:

(1) Lushene (1967) A theory derived from the entire 566 items MMPI which was administered to 189 male undergraduate college students. 18 factors were postulated.
(2) Barker, Fowler, and Peterson (1971) A 373 item short form of the MMPI administered to 1575 VA male hospital patients comprised the source of the 9 factor theory.
(3) Tryon (1966) The entire 566 item MMPI, administered to 310 adult subjects, sex not specified, was the origin of a 7 factor theory. Tryon performed an oblique factor rotation. The first three factors were only slightly correlated, whereas the remaining four are moderately correlated with the first three; therefore the present treatment of his factors as orthogonal should be regarded as suggestive only.
(4) Conventional theory The entire MMPI (511 items at the time) was presumably involved in selecting items for the 9 standard clinical scales, which represent Kraepelinian nosological groupings in vogue at the time of construction of the scales. The remaining items are treated as filler items.

METHOD

DATA

Item answers of 225 male subjects to the original 511 items of the MMPI comprised the data for the study. This is the original sample of males used to determine the standard male norms for the MMPI scales.

COMPUTER PROGRAM

Horst's indirect factor method was translated into a Fortran Computer Program by Barker (1973). The program has been subsequently refined by Stallings (1973) and by Sloan (1973). The computer program clusters variables into a priori subsets, totals the variables in the subsets, determines principal components of the totals and then estimates the principal axes loads of individual variables. The principal axis solution is finally rotated to a varimax criterion. A factor load of + or - .3 or greater on only one factor is required to identify a variable with a factor.

COMPUTER

The computer program was originally developed for an IBM 360 Mod 50 Computer, with 120 K core storage allotted. It was necessary to make extensive alterations in the program to enable it to run accurately on a Univac 1110 system with 128 K core storage allotment. The program capacity is approximately 2000 variables which can be factored.

D MEASURE

The information measure D (relative uncertainty reduction) is employed to relate the degree of association between theorized factor structure (specified subsets of items) and actually obtained factor structure.

RESULTS

For each of the competing theories respectively, Tables 1, 2, 3, and 4 provide a visual display of the association between theorized and obtained factor structure. If the agreement between theory and obtained factor structure were perfect, the diagonal cells of the table would be filled and the off-diagonal cells would be empty. The D measure would then reflect perfect agreement with a value of 1.00. To the extent that cell entries appear in off-diagonal cells, a lack of agreement between theorized and obtained factor structure is indicated. The limit of disagreement results in an equal number of entries in each cell and the resulting D measure is 0.0.

The D values computed for each of the theories are rank ordered from highest to lowest as follows:

(1) Lushene D = .56
(2) Barker, Fowler, and Peterson D = .52
(3) Tryon D = .42
(4) Conventional D = .42

It is clear that none of the four theories provides a satisfactory theory of the factor structure of the MMPI in this population of subjects. The Lushene and Barker et al theories are virtually identical in D value whereas the Tryon and conventional theories are clearly inferior. The Barker et al theory is more parsimonious than the Lushene theory in that only 9 factors are specified whereas Lushene's theory specifies 18. The results with respect to Tryon's theory should be regarded as tentative. Tryon theorized oblique factors whereas the method presently employed treats all factors as orthogonal.

DISCUSSION

The size of the obtained D values relating the theorized and obtained factor structures indicates that none of the four theories is adequate in representing the dimensionality of this data set. Several lines of conjecture merit consideration.

First, none of the existing theories may be truly adequate. This could conceivably result from defects in methodology and/or procedure in arriving at the competing theories.

Second, the population of males, used to derive standard male norms for the MMPI, may represent an unusual cross section of individuals, which serves to conceal some factors and high light other or different factor dimensions. Perhaps it should be noted here that the male subjects were relatives of neuropsychiatric patients, whom they were visiting in the hospital.

Third, the size of the sample used for norming purposes must be viewed as grossly inadequate by current multivariate research standards. For example, Nunnaly (1967) argues for at least ten subjects per variable while Cattell (1966) views a desirable sample size as 100 + the number of variables. The number of subjects in this study falls far short of either Nunnaly's or Cattell's criterion. Thus, instability of findings due to the unfavorable ratio of number of variables to number of subjects may have produced inaccurate results.

Another interesting matter relates to the concept of false positives. The term accurately relates to three of the theories Lushene, Tryon, and conventional theory. Each of these three theories was developed from the entire MMPI. Thus, the appearance of items on factors which were not predicted by the theory are accurately labeled false positives. But in the case of the Barker, Fowler, and Peterson theory, the theory was derived from a subset of the entire MMPI (Viz. 373 items). Therefore the 9 factor theory did not make prediction for items not included in the 373 item set. Identification of these items as false positives is actually inappropriate. Nevertheless the items are so regarded for this study.

A possible solution of this issue is to extend the nine factor theory to encompass the new items (false positive items not included in the original 373 items on which the theory was derived). The new items thus identified with the theory, the indirect factor method would again be employed and the resulting solution evaluated by the D measure. Of course the altered 9 factor theory would now be viewed as a revision of the earlier theory. This general procedure of revising theorized factor structure on the basis of false positives (and perhaps also false negatives) is currently under investigation as a procedure to:

(1) improve the accuracy of theories regarding factor structure.
(2) develop factor theory in areas where none presently exists.

SUMMARY

An indirect method of factor analysis was used to evaluate four competing theories relating to the factor structure of the MMPI for the male population on whom the original MMPI standardization was performed.

Results clearly indicated the inadequacy of each of the four theories and provided a rank ordering of their relative accuracy.

REFERENCES

Barker (Hunnicutt), B. M., and Barker, H. R. Application and evaluation of a novel and indirect factor method to a very large data matrix. Paper presented at the Southeastern Psychological Association, Hollywood, Florida, May 3, 1974.

Barker, B. M., and Barker, H. R. Search for a method of increasing the accuracy of a novel, indirect method of factor analysis. Paper presented at Southeastern Psychological Association, Atlanta, Georgia, March 28, 1975B.

Barker, B. M., Barker, H. R., and Carlton, B. B. Accuracy of an indirect factor analytic solution as a function of the method of assigning variables to totals. Proceedings of the Social Statistics Section, American Statistical Association, 1975, 295-297.

Barker, H. R. Behavioral Sciences Statistics Program Library. University of Alabama, 1973.

Barker, H. R., and Barker, B. M. Accuracy of an indirect factor analytic solution as a function of the number of totals used. Paper presented at the Southeastern Psychological Association, Atlanta, Georgia March 28, 1975A.

Barker, H. R., and Barker, B. M. An indirect factor method for testing the dimensionality of large data sets. Paper presented at Southeastern Psychological Association, New Orleans, Louisiana, March 19, 1976.

Barker, H. R., and Barker, B. M. Summary of Research on a novel, indirect factor analytic solution. Proceedings of the Social Statistics Section, American Statistical Association, 1975C, 298-301.

Barker, H. R., Fowler, R. D., and Peterson, L. D. Factor analytic structure of the short form MMPI items in a VA Hospital population. Journal of Clinical Psychology, 1971, Vol. XXVII, No. 2, 228-233.

Cattell, R. B. (Ed.). Handbook of multivariate experimental psychology. Chicago: Rand-McNally, 1966.

Hamlett, C. C. Validation of the theorized factor structure of the personal orientation inventory. Paper presented at American Statistical Association, Boston, Massachusetts, August 23, 1976.

Hathaway, S. R., and McKinley, J. C. MMPI manual, Revised Ed. New York: Psychology Corporation, 1951.

Horst, P. Factor analysis of data matrices. New York: Holt, 1965.

Lushene, R. E. Factor structure of the MMPI item pool. Unpublished master's thesis, Florida State University, 1967.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Sloan, H. C. A simplified procedure for estimating the factor structure of large data matrices. Unpublished doctoral dissertation. University of Alabama, 1973.

Stallings, N. A. Evaluation of an indirect method of estimating the factor structure of large data matrices. Unpublished doctoral dissertation, University of Alabama, 1973.

Tryon, R. C. Unrestricted cluster and factor analysis, with applications to the MMPI and Holzinger-Harman problems. Multivariate Behavioral Research, 1966, Vol. 1, 229-244.

TABLE 1

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (CONVENTIONAL THEORY)

ITEM SUBSETS X	Factors (Y) I	II	III	IV	V	VI	VII	VIII	IX	False -	Sum
1	6									2	8
2		4			1					18	23
3		5	2							9	16
4		2		4		1	1			13	21
5		4			3		2			12	21
6		7				6				4	17
7		10					1			7	18
8		15		1						22	38
9										19	19
False +	8	53					6				67
Sum	14	100	2	5	4	7	10	0	0	106	248

H(X) = 1.884 H(X,Y) = 4.184
H(Y) = 3.085 HT = .785 D = .42

TABLE 2

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (TRYON THEORY)

ITEM SUB-SETS X	Factors (Y) I	II	III	IV	V	VI	VII	False -	Sum
1	19							6	25
2		20						12	32
3			12					13	25
4			1	10				16	27
5			2		1			17	20
6			5	1		6		9	21
7			1	9			1	23	34
False+		5	6	1		2	1		15
Sum	19	25	27	21	1	8	2	96	199

H(X) = 2.232 H(X,Y) = 4.255
H(Y) = 2.959 HT = .936 D = .42

TABLE 3

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (BARKER, FOWLER AND PETERSON THEORY)

ITEM SUBSETS X	Factors (Y) I	II	III	IV	V	VI	VII	VIII	IX	False –	*Sum
1	*20						8			14	* 42
2	*	23					1			25	* 49
3	* 2		14							11	* 27
4	*			11						11	* 22
5	*		1		9					4	* 14
6	*10						1			7	* 18
7	*						22			11	* 33
8	*							11		1	* 12
9	*						2		4	2	* 8
False+	30	3	3	1	0	0	22	2			* 61
Sum	*62	26	18	12	9		56	13	4	86	*286

H(X) = 2.664 H(X,Y) = 4.363
H(Y) = 3.084 HT = 1.385 D = .52

TABLE 4

ASSOCIATION BETWEEN ITEM SUBSETS AND VARIMAX FACTORS (LUSHENE THEORY)

ITEM SUBSETS X	Factors (Y) I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	False –	Sum
1	31																		17	48
2	8																		15	23
3	1		7								1								10	19
4	3			10									1						10	24
5	2																		14	16
6						2													8	10
7							1													1
8								8	1										4	13
9	1					4													6	11
10						6													4	10
11						2					3								5	10
12												1							5	6
13										2			2					1	2	7
14														2					6	8
15													1		5				2	8
16											2					3			3	8
17								1									4		2	7
18	1																	3		4
False +	54										3	1	16	1				1		76
Sum	101		7	10		14	1	9	1	2	9	2	20	3	5	3	4	5	113	309

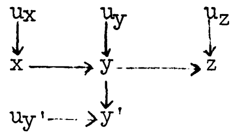
H(X) = 2.648 H(X,Y) = 4.832
H(Y) = 3.668 HT = 1.484 D = .56

CONSEQUENCES OF NON-RANDOM MEASUREMENT ERROR IN PERCENTAGE DIFFERENCE TABLES

Henry Jay Becker
Johns Hopkins University

In the social sciences, attention to the precise measurement of individual variables on individual cases has been supported in principle far more than in practice. However, in several of the disciplines efforts have been made both to educate other researchers of the biasing effects of measurement error and to improve the way in which necessarily "soft" data may be employed in analysis. There have been several approaches taken to the problem reflecting the different academic disciplines and corresponding traditions of research methodology involved.

Sociologists involved in the complexities of multivariate causal models have begun employing both measured but fallible variables and their corresponding unmeasured but "true" variables in their analytic efforts. The causal modeling approach to measurement error is typified by the following causal diagram which postulates measurement error for variable y only, measurement error which is random with respect to x and z, and no direct causal relationships between variables x and z.



These assumptions allow one to estimate the causal paths p_{yx} and p_{zy} (equivalent in this case to the "true" correlation coefficients) from the observed correlations $r_{xy'}$, $r_{zy'}$ and r_{xz}. (Heise, 1969)

In fact, $p_{yx} = \sqrt{\frac{r_{xz'}\ r_{xy'}}{r_{y'z}}}$ and $p_{zy} = \sqrt{\frac{r_{y'z}\ r_{xz}}{r_{xy'}}}$.

For example, suppose $r_{xy'} = .2$, and $r_{y'z} = .4$ and $r_{xz} = .1$. If the above causal model is assumed, then $p_{yx} = r_{xy} = .22$ and $p_{zy} = r_{yz} = .45$. Also $r_{yy'} = .89$.

Most work by Sociologists in this area has followed this approach or a companion "multiple indicators" approach (Hauser and Goldberger, 1971). The research for the most part has been limited to situations where measurement error for a given variable can be assumed to be randomly distributed with respect to other variables in the causal system and other error terms as well. (For an exception, see Sullivan, 1974.) This is primarily because these methodologists, dealing with many variables simultaneously, have had their hands full just incorporating random measurement error effect into their measurements. In the end, however, problems produced by non-random measurement error will necessarily have to be attended to.

The consequences of non-random measurement error have been addressed by the economists, Lansing and Morgan (1971). They have shown, for simple bivariate correlations between continuous variables, the consequences of a number of different types of non-random measurement error. Measurement error is discussed in the context of evaluating its effects on the regression coefficient $\beta_{y'x'}$ where the primes denote observed variables. If the observed values are composed of a "true" value plus an "error" term (e.g., $y' = y + v$), the estimating equation for β in terms of "true" scores and "error" terms becomes:

$$\beta_{y'x'} = \frac{\text{cov } xy + \text{cov } xv + \text{cov } yu + \text{cov } uv}{\text{var } x + 2\text{cov } xu + \text{var } u} \quad (1)$$

and that for the correlation coefficient:

$$r_{x'y'} = \frac{\text{cov } xy + \text{cov } xv + \text{cov } yu + \text{cov } uv}{\sqrt{\text{var } x + 2\text{cov } xu + \text{var } u}\ \sqrt{\text{var } y + 2\text{cov } yv + \text{var } v}}. \quad (2)$$

An examination of equations (1) and (2) shows the role played by three types of non-random measurement error as well as the role of random measurement error:

(a) Association between the error deviate terms u and v. (cov uv) Positive association between error terms spuriously raises the reported association between x and y. Negative assocation between error terms attenuates the reported x:y relationship from the true value. A large enough negative assocation between error terms could reverse the reported association between x and y.

(b) Association between the true value of one variable and the error term of the other. (cov xv + cov yu) This has the same kinds of effects as (a); e.g., positive associations spuriously raise correlation and regression coefficients while negative associations reduce them or reverse them.

(c) Association between the true value of one variable and its error deviate. (cov xu; cov yv) With the placement of these factors in the denominator of Eq. (1) and (2), a positive association here would lower the β or r term from its true value while a negative association would increase the values of these statistics.

(d) Random error (var u; var v) Since these terms, also found in the denominator, can only be positive, their presence always causes regression and correlation coefficients to be underestimated.

Categorical variables can also be examined in the above framework. Consider the association between two dichotomies each coded "0" and "1." In this case, the true value is always negatively correlated with its own error term, since a true "0" score will have an error term which is either "0" (accurately measured true score) or "+1" (measured in error) while a true "1" score will have a "-1" error term when falsely reported as "0." Alone, this factor would cause reported associations to be higher than the true association. At the same time, however, the magnitude of the errors--or more precisely, the total variance of the errors of observation in each variable considered separately--is acting to spuriously lower the reported association from the true value. This latter factor is universally more powerful, as can be shown by extensive algebraic manipulation.

Thus, whenever there is no association between the error term of each variable and the true score of the other and where error terms themselves are uncorrelated, the net effect of the other two factors is to spuriously reduce measured association from its true value.

Conversely, wherever error terms are sufficiently positively correlated or where there is a large enough positive correlation between true values of one variable and error terms of the other, these factors can overcome the general tendency for attenuation of relationships. In particular, for the regression coefficient to be spuriously high, the offending factors (cov xv + cov yu + cov uv) must be β_{yx} times as great as the factors tending to diminish the measured relationship (cov xu + var u). The inequality with respect to the correlation coefficient is somewhat more complex: $(\text{cov } xv + \text{cov } yu + \text{cov } uv)^2 > r_{xy}^2 [\text{var } y (\text{cov } xu + \text{var } u) + \text{var } x (\text{cov } yu + \text{var } v) + (\text{cov } xu + \text{var } u)(\text{cov } yv + \text{var } v)]$.

One of the principal advantages of working with dichotomies and categorical variables in general is the clarity with which statistical phenomena can be exemplified. To examine the effects of non-random measurement error on reported associations between variables, consider the data in Tables 1a and 1b.

Each table shows both the true and reported values of a variable X scored "+" and "-" cross-tabulated by a variable Y, whose "types" represent population subgroups. Variable Y is assumed to be know without error. For each "type" (i.e. population subgroup), 90% of the x = "+" cases are known without error to be "+." (The other 10% are called "false negatives.") For two types (B and C), 95% of the x = "-" cases are known without error; only 5% are "false positives." In the other two types (A and D), false positives number 20% of the true x = "-" cases. Overall, more than 88% of the cases are accurately scored on both variables X and Y.

However, in neither case is the result similar to what it would have been if only random measurement error (i.e. equal error rates) were present. If, for example, the error rate were 10% in each cell, the reported percent difference would have been 16% instead of 20%, making the reported result a "conservative" estimate of the true relationship. Instead, in Table 1a, the reported association slightly exaggerates the true association between the variables. (In percentage difference terms, 20% points true difference is increased to 26%.) In Table 1b, the association is markedly understated by the reported data (the 20% true difference is reduced to 5%).

Several earlier papers by biostatisticians involved in public health research have discussed the

TABLE 1: EFFECTS OF DIFFERENTIAL FALSE POSITIVE RATES ON REPORTED ASSOCIATIONS (HYPOTHETICAL DATA)

Table 1a

TRUE VALUES

	Type A	Type B
+	40%	20%
-	60%	80%
Total	(100)	(100)

%difference = 20 percentage points
gamma = .45
odds ratio = 2.7

ERROR RATES

True Values	Type A	Type B
+(FN)	10%	10%
-(FP)	20%	5%

REPORTED VALUES

	Type A	Type B
+	48%	22%
-	52%	78%
Total	(100)	(100)

% difference = 26 percentage points
gamma = .53
odds ratio = 3.3

Table 1b

TRUE VALUES

	Type C	Type D
+	40%	20%
-	60%	80%
Total	(100)	(100)

% difference = 20 percentage points
gamma = .45
odds ratio = 2.7

ERROR RATES

True Values	Type C	Type D
+(FN)	10%	10%
-(FP)	5%	20%

REPORTED VALUES

	Type C	Type D
+	39%	34%
-	61%	66%
Total	(100)	(100)

% difference = 5 percentage points
gamma = .11
odds ratio = 1.2

issue of non-random error's effects on measured association using cross-classified dichotomies, but their work needs more widespread attention and follow-up.

Keys and Kihlberg (1963) graphed the expected bias on individual variables (true vs. reported "prevalences"--i.e. percent "+") for a variety of false positive and false negative rates. They also graphed the bias of an associational measure (proportionate deviation from a true "relative risk" of 1.0) against that common true risk given particular combinations of the four error rates involved. However, the authors did not calculate bias in the measured assocation under the condition that different (true) rates were present (i.e., true relative risk $\neq 1.0$). They did explore the complexities of calculating estimated bias with fallibility on both measured variables, but produced no graphical examples as in the case of a single variable measured in error. (Instead of only four parameters--two false positive and two false negative rates--that situation involves 16 parameters.)

Goldberg (1975), examining the same type of hypothetical data, found that subgroup differences in false positive rates produces far larger average bias than differences between false negative rates for "prevalences" under 50%; coversely for higher prevalences.

One other point shown in the above example is that the direction of bias depends on whether the group with the greater or the lesser proportion of "+" cases is the one with the larger proportion of false positives. If false positives are more frequent in the group with more "+" cases, generally the reported relationships will be an exaggeration of the true relationship. Conversely major underreports of association tend to occur where false positives are concentrated in the group with fewer "+" cases. These results are reversed and apply to false negative rates where "+" proportions are above 50%.

In general, the reported percentage difference is related to the true percentage difference and the various error factors as $(p_1'-p_2') = p_1(1-FN_1 - FP_1) - p_2(1-FN_2 - FP_2) + (FP_1 - FP_2)$ where for subgroups i = 1, 2, p_i' are reported proportions "+," p_i are true proportions "+", FN_i are false negative rates (proportions of true "+" reported as "-") and FP_i are false positive rates (proportion of true "-" reported as "+"). Each p_i is also related to the reported p_i' and the error terms, thusly:

$$p_i = \frac{p_i' - FP_i}{1-FN_i-FP_i}.$$

Under the assumption that all error rates, FN_i and $FP_i \leq e$, where $p_1 - p_2 \geq 0$, the maximum and minimum value for $(p_1 - p_2)$ are given by

$$\left(\frac{1}{1-e}\right)\left(p_1' - p_2'\right) \pm \left(\frac{e}{1-e}\right).$$

As shown in Table 2, extremely large variations in the magnitude and direction of bias can be observed due to non-random measurement error even when the absolute proportions of cases in error is rather modest, for example e = .2.

In most cases, although not for the maxima and minima shown in Table 2, the magnitudes of p_1' and p_2' as well as their difference $(p_1' - p_2')$ affect the value expected for $(p_1 - p_2)$ under specific error rates. (See Table 3.) The smaller the magnitude of p_1' and p_2' (given $p_1' - p_2' = k$), the greater the difference between true and reported associations.

TABLE 2: RANGES OF POSSIBLE TRUE PERCENTAGE DIFFERENCES GIVEN CERTAIN LIMITS ON MEASUREMENT ERROR

REPORTED VALUES[2] $(p_1'-p_2')$	MAX & MIN TRUE VALUES $(p_1 - p_2)$ FOR MAXIMUM ERROR RATE e[1]					
	e=.3		e=.2		e=.1	
	MAX	MIN	MAX	MIN	MAX	MIN
+.3	.86	.00	.62	.12	.44	.22
+.2	.71	-.14	.50	.00	.33	.11
+.1	.57	-.29	.38	-.12	.22	.00
+.05	.50	-.36	.31	-.19	.17	-.06
.00	.43	-.43	+.25	-.25	.11	-.11

1. excluding effects of sampling error
2. with $p_2' > p_1'$, results are the same except minima and maxima are reversed and signs are reversed also.

Even rather minor differences in the two false positive rates can produce rather major differences in measured association. The example in Table 4 shows that under rather ordinary conditions, a reversal of false positive rates changes the result from nearly no measurement error effect (situation (1)) to a very strong attenuation of the true relationship (situation (2)).

The important questions that are raised by this discussion of non-random error concern not the possible effects of differential error rates, but the actual effects caused by real differences in error rates. Here our knowledge is hampered by the paucity of validation studies that report validity for subgroups rather than solely for the complete sample studied. Two sources have been found in the public opinion literature which do report on validity of responses for subgroups.

One is the Denver validity study done in 1949 but most recently analyzed in (Cahalan, 1969). Reworking the results on the basis of the published statistics, interesting and quite major differences between subgroups can be seen for both false negative and false positive rates, with the consequence that many relationships would have been reported erroneously without the validating information.

ABLE 3: EFFECTS OF MAGNITUDE OF p_i' AND $(p_1' - p_2')$ ON BIAS OF $(p_1'-p_2')$ GIVEN FOLLOWING ERROR RATES: $FN_1 = .10$ $FN_2 = .10$ $FP_1 = .20$ $FP_2 = .10$

REPORTED VALUES			TRUE VALUES	BIAS
$p_1'-p_2'$	p_1'	p_2'	p_1-p_2	$(p_1-p_2)-(p_1'-p_2')$
.30	.5 .4	.2 .1	.30 .29	.00 -.01
.20	.5 .3	.3 .1	.18 .14	-.02 -.06
.10	.5 .3	.4 .2	.05 .02	-.05 -.08
.05	.5 .3	.45 .25	-.01 -.05	-.06 -.10
.00	.5 .3	.5 .3	-.07 -.11	-.07 -.11
-.05	.45 .25	.5 .3	-.14 -.18	-.09 -.13
-.10	.4 .2	.5 .3	-.21 -.25	-.11 -.15
-.20	.3	.5	-.36	-.16
-.30	.2	.5	-.50	-.20

TABLE 4: EFFECTS OF SMALL DIFFERENCES IN FALSE POSITIVE RATES

TRUE SCORES

	A	B
%+	25%	15%

% difference = 10 percentage points
gamma = .31
odds ratio = 1.9

ERROR RATES (1)

	A	B
FN	12%	12%
FP	15%	10%

ERROR RATES (2)

	A	B
FN	12%	12%
FP	10%	15%

REPORTED RESULTS (1)

	A	B
%+	33%	22%

% difference = 11 percentage points
gamma = .27
odds ratio = 1.7

REPORTED RESULTS (2)

	A	B
%+	30%	26%

% difference = 4 percentage points
gamma = .10
odds ratio = 1.2

The Cahalan data is only approximately reproduced in Table 5. (The absence in the original article of sufficient information about the magnitude of "don't knows" for men and women separately made it impossible to exactly reproduce the data.) The data as estimated here, though, exhibit some remarkable differences between cells in error rates for some of the measured associations. Notice that in both cases of large reported associations by sex (the Community Chest contributions and the drivers' licenses), the true relationships are of smaller magnitude--one of them being less than 40% as large as the reported association.

The second article from Public Opinion Quarterly reporting validity coefficients for subgroups is Weiss's (1968) report of interviews with mothers receiving welfare. The results given in Table 6 illustrate the tendency for upward bias in reported percentage differences to occur only where both $p_1 - p_2$ is small and differences in false positive rates exist. For cases where $p_1 = p_2$, bias is also greater when the p_i approach zero. (See Keys and Kihlberg, 1963.)

Finally, the question remains: given the clearly major consequences of non-random measurement error on reported associations between variables, what now needs to be done to better take account of this problem?

Journal articles most often report findings with due respect for sampling error. Confidence limits, for example, are reported for percentage differences and the result is reported as statistically significant if the null hypothesis of "no difference" lies outside the range of 2 standard errors of the sample statistic. It would seem to make some sense to include in this statement the range of uncertainty that is attributable to possible measurement errors.

That is, the analyst might propose two sets of error rates for a given two-by-two table that would have opposite effects on the bias of the reported result and which are just extreme enough to be plausible given the possible error-producing causes present in the particular data. Those two sets of error rates, in turn, would be applied to the reported data in order to produce "extreme" but plausible "true" sets of sample data. Confidence intervals would then be calculated for these extreme but hypothetically true sample data tables. The summary confidence intervals reported would thus reflect both sampling error and plausible ranges of error caused by measurement inadequacies.

Consider an example, more or less randomly chosen from Rosenberg's The Logic of Survey Analysis (1968). (See Table 7.) In our use of his re-examination of the relationship between vote intention and respondent's education (1948 data), we percentage on education (% with some high school) taking education as the variable possibly reported in error. Vote intention is assumed to be reliably and validly known. Let us make two contrasting suppositions: (t_1) that persons claiming an intention to vote are more

TABLE 5: MEASUREMENT ERROR EFFECTS IN THE DENVER VALIDITY STUDY (1949)

	(N)	REPORTED[1]	TRUE	ERROR RATES[2] FN	FP	Gammas
Community Chest Contributions (percent "yes")						
MEN	327	85%	35%	0%[3]	76%	r=.65
WOMEN	404	55%	24%	0%	41%	t=.26
△		30%	11%			
Driver's License (percent "yes")						
MEN	423	78%	66%	3%	41%	r=.76
WOMEN	497	33%	32%	12%[4]	8%	t=.61
△		45%	34%			
Voting in 1946 Cong. Elections (percent "yes")						
MEN	382	55%	32%	5%	37%	r=.14
WOMEN	447	48%	33%	9%	27%	t=-.02
△		7%	-1%			
Voting in 1948 Pres. Elections (percent "yes")						
MEN	423	74%	61%	1.6%	36%	r=.05
WOMEN	497	72%	61%	1.3%	31%	t=.00
△		2%	0%			

1. Percentage recalculated from published data omitting "don't knows."
2. FN = false negative FP = false positive
3. Not checked: assumed = 0.
4. Unable to reproduce marginals from published data: doubtful accuracy.

TABLE 6: MEASUREMENT ERROR EFFECTS IN THE WEISS (1968) STUDY OF WELFARE MOTHERS

PERCENT VOTING IN 1964 ELECTION						
	(N)	REPORTED	TRUE	ERROR RATES[3] FN	FP	Gammas
Worked more than 10 years[1]	189	53%	29%	0%[2]	34%	r=.17
Worked less than 10 years	329	44%	28%	0%	22%	t=.02
△		9%	1%			
Expects children to continue educ. past high school	141	64%	42%	0%[2]	39%	r=.33
Does not expect...	191	47%	29%	0%	26%	t=.29
△		17%	13%			
Did not get as much education as wanted	372	48%	28%	0%[2]	27%	r=.02
Got enough education	147	47%	28%	0%	26%	t=.00
△		1%	0%			

1. Assumed known without error
2. Assumed to be zero
3. FN = false negative FP = false positive

likely to overreport their education than those not claiming an intention to vote (because of the good citizen image of voting and completing one's education); (t_2) alternatively, that persons not planning to vote, being more alienated from the culture, make more errors in their self-reports (both over-reporting and underreporting) than do those planning to vote. Table 7 shows the the four parameters assumed for each of these two basic models. The data in Table 7 show how the reported relationship (a) is affected by the measurement error assumptions (t_1) and (t_2) producing hypothesized "true" sample results (b) and (c). If the confidence intervals for (p_1-p_2) are calculated from the "extreme" results (b) and (c) instead of from (a), they change the estimate of (p_1-p_2) from $8\% \leq (p_1-p_2) \leq 20\%$, $p = .95$, to $5\% \leq (p_1-p_2) \leq 24\%$, $p = .95$.

TABLE 7: ILLUSTRATION OF THE USE OF MEASUREMENT ERROR HYPOTHESES IN THE ROUTINE PRESENTATION OF CONFIDENCE INTERVALS[1]

	REPORTED EDUCATION: % WITH SOME HIGH SCHOOL (% "+")		EXPECTED UNDER ERROR HYPOTHESES	
			t_1	t_2
Vote Intention:	(a)	(N)	(b)	(c)
Positive	59%	(2515)	52%	58%
Negative	43%	(297)	41%	40%
△	16%		11%	18%
Confidence Intervals:				
p=.95	$10\% \leq (p_1 - p_2) \leq 22\%$		$5\% \leq (p_1 - p_2) \leq 24\%$	
p=.68	$13\% \leq (p_1 - p_2) \leq 19\%$		$8\% \leq (p_1 - p_2) \leq 21\%$	

t_1		Education FN	Education FP
Vote Int:	Pos	.01	.15
	Neg	.03	.05

t_2		Education FN	Education FP
Vote Int:	Pos	.02	.05
	Neg	.07	.10

1. Data from Rosenberg, 1968

One standard error confidence limits for the data based on the measurement error assumptions do not differ much from the two-standard error limits based solely on sampling error: $8\% \leq (p_1 - p_2) \leq 21\%$.

Another procedure that could be more widely utilized is a version of Schuman's suggested "random probe" method of clarifying the meaning of information obtained through interview procedures. Schuman (1966) proposed that respondents to a given survey be asked to give reasons, interpretations, or explanations for their answer to a subset of all closed-ended questions in the interview. These extended responses are used to calculate the proportion of closed-ended responses that were "accurate." A different subset of respondents is used to "validate" in this way each of several subsets of questions.

This random probe technique has the potential of providing information concerning response error rates for various subgroups being compared in the statistical analysis. Thus real data can be employed as "expected values" in the error rate table, and variances for these error rates can be computed. These variances can be used to produce the "extreme" values of the "true" results which can then be used, along with the sampling distribution of the statistic, to compute reasonable confidence limits for the data.

Obviously, much additional work needs to be done in order to make it possible for measurement error effects to be incorporated into routine statistical procedures to the extent that sampling error has been. But the expanding application of complex statistical methodologies to the "soft" data of the social sciences has not been accompanied by sufficient appreciation of the magnitude and direction of effects caused by unreliability and invalidity of measurement. This is a problem that requires major attention if we are to avoid perpetuating the publication of technically flawless conclusions invalidated by the use of extremely soft and fallible data.

References

Cahalan, Don, "Correlates of Respondent Accuracy in the Denver Validity Survey," Public Opinion Quarterly, Vol. 32, (Winter, 1968), 607-621.

Goldberg, Judith D., "The Effects of Misclassification on the Bias in the Difference Between Two Proportions and the Relative Odds in the Fourfold Table," Journal of American Statistical Association, Vol. 70 (September, 1975), 561-567.

Hauser, Robert M. and Goldberger, Arthur S., "The Treatment of Unobserved Variables in Path Analysis," in Sociological Methodology 1971, Herbert L. Costner, ed. Jossey-Bass, 1971.

Heise, David R., "Problems in Path Analysis and Causal Inferences," in Sociological Methodology 1969, Edgar F. Borgatta, ed. Jossey-Bass, 1969.

Keys, Ancel and Kihlberg, Jaakko K., "Effect of Misclassification on Estimated Relative Prevalence of a Characteristic," American Journal of Public Health, Vol. 53 (October 1963), 1656-1665.

Lansing, John B. and Morgan, James N., Economic Survey Methods. Institute for Social Research, University of Michigan, 1971.

Rosenberg, Morris, The Logic of Survey Analysis. Basic Books, 1968.

Schuman, Howard, "The Random Probe: A Technique for Evaluating the Validity of Closed Questions," American Sociological Review, Vol. 31 (1966), 218-222.

Sullivan, John L., "Multiple Indicators: Some Criteria of Selection," in Measurement in the Social Sciences, H.M. Blalock, Jr., ed. Aldine, 1974.

Weiss, Carol H., "Validity of Welfare Mothers' Interview Responses," Public Opinion Quarterly, Vol. 32 (Winter, 1968), 622-633.

LIFE TABLE METHOD WITH PARTIAL AND INCOMPLETE FOLLOW-UP IN A STUDY ON REEDUCATION OF DWI DRIVERS

James Beckett III, Bowling Green State University
Wayne A. Woodward, Southern Methodist University

INTRODUCTION AND SETTING

Because of the growing concern for the problem of alcohol-related traffic accidents, the Department of Transportation developed Alcohol Safety Action Programs (ASAP's) in selected communities across the United States. An ASAP project was implemented in Dallas, Texas in October 1972. A major goal of this particular program was to develop, investigate, and evaluate methods of educating the DWI (driving while intoxicated) driver. During the 27 month period from the inception of the study until December 1974, 6644 persons were arrested on first offense DWI charges and assigned to the program as a condition of their probation. These probationers were randomly assigned to the control group or to the "school" group, where the school group was composed of three specific models. Thus members of the school group were further apportioned to either the Traditional Model (TM), Counseling Model (CM), or the Parent Effectiveness Training (PET) Model.

The purpose of the Traditional Model was to provide information to the probationers concerning alcohol, its effects on the body, and its effects on driving. The Parent Effectiveness Training Model was designed to teach interpersonal communications skills in the areas of conflict resolution. It was felt that these skills would equip the person to deal with problems through means other than drinking. The Counseling Model was conceived relatively late in the program and was based on the idea that "problem drinkers" and "social drinkers" should be treated differently. Probationers were classified as either a social drinker or a problem drinker on the basis of the Mortimer-Filkins Alcohol Abuse Inventory. Social drinkers were then assigned to classes identical to the TM classes. The problem drinkers attended sessions in which dyadic interaction between probationers along with brief direction from instructors was employed to help each problem drinker identify his specific alcohol-related problem.

STATISTICAL ANALYSIS OF MODEL PERFORMANCE

The objective of the previously described models was to eliminate (or reduce) the drinking/driving behavior of the probationers. The effectiveness of each model was measured by the rearrest rate of its graduates. Thus the three models and the control group were compared by monitoring the DWI arrest data for the city of Dallas in order to obtain data concerning which probationers were arrested on a DWI charge after their training (or assignment to the control group), and for those who were rearrested, how long they had been out of school when they were rearrested.

Due to the magnitude of the study, a probationer was observed in this fashion only if he was rearrested on a DWI charge in the city of Dallas. With this partial individual follow-up it was impossible to determine whether or not a probationer had been rearrested somewhere other than Dallas or even whether or not he still lived in Dallas. Those graduating from a school in the latter part of the study period (e.g., Counseling Model) were less exposed to the possibility of rearrest.

Biostatisticians handle a similar problem in medical research and epidemiological studies using the Actuarial or Life Table Method. For those unfamiliar with the method, Cutler and Ederer (1958) provide a well-written example of the procedure. In the medical setting measurement of patient survival is necessary for the evaluation of the treatment of usually fatal chronic diseases, e.g., cancer. The principal advantage of the Life Table Method (here) is that it makes possible the use of all survival (rearrest) information accumulated up to the closing date of the study. However, cancer patients are usually followed-up periodically until the end of the study. Our situation with partial and incomplete follow-up magnifies the usual problem of competing risks. The "partial and incomplete" follow-up stems from the fact that only those few who are rearrested are part of the follow-up; the many others not rearrested were never contacted in any way. If a person leaves the study (moves from Dallas or quits driving in Dallas), this information should be used in estimating the rearrest rate. In our case this outward migration was estimated (from outside sources) to be 10 percent per year which causes the rearrest rates to be adjusted slightly upward. Withdrawals from the study and losses due to follow-up (migration) were assumed uniformly distributed throughout each quarter.

Since the duration of the study was only nine periods (quarters), the effect of 10 percent annual migration versus, say 8 percent or 12 percent is not as pronounced as it would have been for a study with longer follow-up. The question of interest is which school is most effective; for this we need only detect significant differences in proportions. The slope of the rearrest rate curve (vs. time) was not considered; only the final estimated proportion since all the schools' objectives were felt to be more short term than long term.

RESULTS AND REMARKS

This general method (Life Table Method) is by no means new and improvements, such as the Cox Model (Cox, 1974), are to be found in the statistical literature. The Cutler and Ederer version of the Life Table Method as mentioned here is simple to use (See Table 1), and studies have shown it to be quite acceptable. However it has been our experience that the awareness of the Life Table Method is not widespread in the social sciences where it might have impact on studies similar to that in which we participated. Our approaches in evaluating person-quarters exposed and estimating outward migration provided us with the following interpretable results after three quarters.

	Cumulative Rearrest Rate	Standard Error of Rearrest Rate	Effective Number Exposed to Rearrest
CM	.0212	.0080	147
TM	.0469	.0048	1733
Control	.0400	.0054	1105

Note that the Counseling Model was significantly the best as covered more fully in Gottlieb, Pascoe, Woodward, and Beckett (1975).

ACKNOWLEDGEMENT

This research was supported by contract No. DOT AL 71-003 from the Texas Traffic Safety Administration, and the Department of Transportation, National Highway Traffic Safety Administration.

REFERENCES

[1] Cox, D. R. (1972). "Regression Models and Life-Tables (with discussion)," JRSSB, 34, 187-220.

[2] Cutler, J. J. and Ederer, F. (1958). "Maximum Utilization of the Life Table Method in Analyzing Survival," Journal of Chronic Diseases, 4, 699-712.

[3] Gottlieb, M. C., Pascoe, E. A., Woodward, W. A., and Beckett, J. (1975). "A Comparison of Treatment Strategies for DWI Offenders," unpublished manuscript.

TABLE 1

Quarters after completion	Number at beginning of period	Number rearrested	Number migration	Number withdrawn	Effective number exposed	Cumulative proportion rearrested
0 - 1	1167	8	29	425	940	.0085
1 - 2	705	3	19	390	501	.0145

THE HISTORY AND FUTURE OF GRAPHICS IN STATISTICS*

James R. Beniger and Dorothy L. Robyn
University of California, Berkeley

The scientific enterprise might be defined as the recasting of individual experience into collective, quantifiable imagery. Throughout the development of science, visual images and numbers have served, side by side, in such forms as coordinate systems and line graphs, histograms and scatterplots. If mathematics reigns as the Queen of Science, then quantitative graphics must rank among her most valuable handmaidens.

Systematic visualization did not begin with modern science, of course, but traces its origins to the dawn of socially-constructed reality. The earliest known map, extant on a clay tablet dated at 3800 B.C., depicts all of Northern Mesopotamia with conventions and symbols still familiar today. From about 3200 B.C., Egyptian surveyors abstracted their lands in terms of coordinates not unlike the Cartesian system still in use. By the 10th century, medieval astronomers depicted planetary movements as cyclic lines on spatial-temporal grids, diagrams strikingly similar to modern line graphs (see 45 for an example).

Despite the central role which collective, quantifiable imagery has played throughout the development of Western thought, there exists only a single monographic treatment of the subject in English (46). This borrows heavily from a French work (48) published nearly a century ago. Except for some renewal of interest in quantitative graphics among historians of statistics (49,50), visual forms remain virtually unnoticed by historians and sociologists of knowledge, in general, and of science, in particular. It is this oversight which the present paper, part of a larger effort in a number of areas (42,43,44), is intended to help correct.

Spatial Organization and Analysis

The problem of spatial organization, of using space to locate--and to assist analyses of-- multiple measurements and multivariate data points, dominated quantitative graphic development during the 17th and 18th centuries. The problem arose with the first stirrings of the Industrial Revolution, in a spate of new measuring devices: the air and water thermometer (c. 1590), micrometer (1636), barometer (1643), pendulum clock (1656), mercury thermometer (1714), etc. Contributions by Descartes, Fermat and other French mathematicians to analytic geometry, plus the reestablishment of Cartesian coordinates in mathematics, provided the basic analytic tools.

In 1686, 43 years after the barometer's invention, and 39 years after Descartes' famous Appendix on analytic geometry (2), Sir Edmund Halley published a Cartesian plot of barometric pressure vs. elevation above sea level(3), the first known analysis of empirical data using a scatterplot.

With such an impressive beginning, it might be reasonable to expect that Cartesian coordinates and analytic geometry became the dominant paradigms for spatial organization and analysis of data by the early 18th century. After an exhaustive search of 18th century journals and texts, however, Tilling concludes that "the use of experimental graphs, either with or without subsequent analysis, never became commonplace in that age" (50, p. 194).

One reason is that coordinate plotting and graphical analysis had major rivals in statistical tables and tabular inference. Tables began to appear in print in the early 1600's. Just as the Cartesian approach came to dominate mathematics, tabular inference captured the social and applied sciences. These alternative approaches to spatial organization and analysis were not reconciled until the 19th century.

Nothing illustrates the obsession of early experimental scientists for tabular data -- and their disregard for graphical plotting and analysis--better than the history of automatic recording devices. Between 1660 and 1800, scores of different mechanical recorders were invented that produced moving line graphs of various natural phenomena: temperature, barometric readings, wind speed and direction, tidal movements, etc. These automatic graphs were considered useless for analysis, and were routinely translated into tabular logs (the literature is reviewed in 47). By the 19th century, after graphical analysis came to be appreciated, scientific journals recorded painstaking graphing and linear extrapolation of data--much of which had originally come from machine-produced line graphs only decades earlier.

The dominance of tabular description was largely due to a vociferous movement of social scientists, which formed under the catchphrase *Die Tabellen-Statistik* in Germany in the early 17th century, and which came to be known as "Political Arithmetic" in Britain after 1687. The movement took impetus from the mounting collection and publication of state statistics or *Staatenkunde*--on population, land and agricultural production -- for the purpose of taxation in the new nation states of Europe.

The eventual triumph of coordinate plotting and graphical analysis is well-documented in the commercial development of graph paper. Rectangular grid paper was first offered for sale by a Dr. Buxton in London in 1794. The first appearance of this paper in published research came in 1800, in an article on barometric variations in *Philosophical Magazine* (24), which included an advertisement for Buxton's product. Herschel made ingenious use of plotted data to calculate the elements of the elliptical orbits of double stars, and his 1832 paper on the subject included a ringing endorsement of graph paper (15, pp. 171-72). Lalanne introduced both a logarithmic grid (20) and polar coordinates (19) to the French Academy in 1843. Twenty years later, Jevons developed the use of semi-logarithmic paper in England (17), and in 1879 he included the first published instructions on the uses of graph paper in the third edition of his *Principles of Science* (18). In 1883, the British government issued the first patent on logarithmic paper.

Graphical methods also received a boost, as the dominant paradigm for spatial organization and analysis of empirical data, from the parallel development of statistical maps, an application for coordinate plotting that could not be duplicated by tabular presentation. In 1701, Halley again took the lead, publishing the first map with empirical

*The first phase of this research, conducted during 1974-75, was supported by Grant GS-29115 from the Division of the Social Sciences, National Science Foundation.

data points, lines of magnetic declination gathered for all navigated waters of the world (5). In 1775, Charpentier published the first geological map, on which he plotted the distribution of various soils and minerals. Seven years later, using similar techniques, Crome published the first of his several geographical plots of demographic, political and economic statistics (1). Minard was the first to plot statistical symbols -- circles, squares and small bar graphs proportional to coal production -- on a map published in 1851 (22).

Discrete Comparison and Continuous Distribution

With the development of descriptive state statistics or Staatenkunde arose the problem of discrete quantitative comparison. Both tables of statistics in general, and comparative political data in particular, suggested the possibility of graphical comparison, especially for the growing volume of atlases and chartbooks intended for mass consumption. The first breakthrough in discrete comparison came in 1765, when Joseph Priestley published the first of his several time-line charts (8), which used individual bars to compare the life-spans of some 2,000 celebrated people.

Priestley's time-lines proved a commercial success and popular sensation, and directly inspired William Playfair's invention of the bar chart, which first appeared in his Commercial and Political Atlas (6), published in 1786. Ironically, Playfair was driven to his invention by lack of data; with the single exception, all of his other plates were line graphs or surface charts, the only published appearance of these forms in the 18th century (50, p. 199).

In the next 15 years, Playfair came to recognize the importance of his contribution. In 1801, he increased the inventory of discrete comparison with

EARLY DEVELOPMENTS IN QUANTITATIVE GRAPHICS

c. 3200 B.C.	Coordinate systems to locate points in real space --Egyptian surveyors
10-11th cent. A.D.	Curves on time grid(planetary orbits) --unknown transcriber of commentary of Macrobius on Cicero's In Somnium Scipionis
12-13th cent.	Musical notation as true time series, with neumes (corresponding to notes) of fixed duration, introduction of bars to mark equal numbers of beats -- Franconian reform, following Franco of Cologne, Ars Cantus Mensurabilis
c. 1350	Proto-bar graph (of theoretical function)--Nicole Oresme (French mathematician)
early 17th cent.	Tables of empirical data -- Die Tabellen Statistik (Germany)
1620	Systematic graphic computation with scale of numbers (forerunner of slide rule, nomography)--Edmund Gunter (Eng. astronomer)
1637	Coordinate system reintroduced in mathematics, analytic geometry, to establish relationship between curve and equation --René Descartes (French)
c. 1660	Automatic recording device (weather-clock, producing moving graph of temperature)--Wren (Eng.)
1686	Bivariate plot of observations (barometric reading vs. height), graphical analysis of empirical data--Edmund Halley(Eng.)
1693	Mortality tables--Halley
1701	Isobar map (lines of magnetic declination, world) --Halley
1712	Literal line graph, inspired by nature of observation (section of hyperbola, formed by capillary action of colored water between two glass plates)--Francis Hauksbee (English)
1724	Abstract line graph (of barometric observations), not analyzed -- Nicolaus Cruquius
1752	Contour map -- Phillippe Buache (French)
1754	Hypothetical mortality curve -- Jean Phillippe Loys de Cheseaux (French)
1760	Curve-fitting and interpolation from empirical data points--J.H. Lambert (German)
1763	Graph of beta density-- Thomas Bayes (English)
1765	Theory of measurement error as deviations from regular graphed line --J. H. Lambert
	Historical time lines (life spans of 2,000 famous people, 1200 B.C. to 1750 A.D.); quantitative comparison by means of bars -- Joseph Priestley (English chemist)
1767-1796	Repeated systematic application of graphical analysis (line graphs applied to empirical measurements)--J.H. Lambert
1774	Graph of density functions -- Pierre Simon de Laplace (Norman)
1775	Geological map (showing distribution of soils, minerals) -- Charpentier (French)
1779	Graphical analysis of periodic variation (in soil temperature)--Lambert
1782	Statistical map --A.W.F. Crome (Professor of Political Economy and Statistics , University of Giessen, Germany)
1785	Superimposed squares to compare areas (of European states) --A.W.F. Crome
1786	Bar chart --William Playfair (English) , Commercial and Political Atlas
1792	Word "chart" for data arranged in graphic or tabular form
1794	Printed coordinate paper -- a Dr. Buxton (English)
1795	Multi - number graphical calculation (contours applied to multiplication table) -- Louis Ezéchiel Pouchet (French manufacturer)
1796	Automatic recording of bivariate data (pressure vs. volume in steam engine), " Watt Indicator " (pressure gauge produced horizontal motion in pen, piston in cylinder added vertical motion) --John Southern and James Watt (English), device kept secret until 1822

two new forms: the pie chart and circle graph (published in 25). Alexander von Humboldt, acknowledging Playfair's influence, combined the bar and pie chart ideas in the subdivided bar graph, first published in 1811 (16).

By the 1820's, a steadily increasing number of the scholarly publications of Europe contained graphs that described and compared (but did not analyze) empirical measurements of a wide range of natural and social phenomena. During the period 1830-35, graphical analysis of natural phenomena finally emerged as a regular feature of scientific publication, particularly in England.

The 1820's and 30's also brought new breakthroughs in the graphical problem of representing continuous distribution. This problem was central to two comparatively new fields: abstract probability theory and vital statistics. Although Halley published the first scientifically-constructed mortality tables in 1693 (4), he apparently never attempted a graphical analysis of his figures. The French mathematicians Loys de Cheseaux and d'Alembert graphed hypothetical curves of mortality, in the mid-18th century, but did not base their curves on actual data—though accurate mortality data had existed for some 60 years.

It appears that the graphics of spatial location lacked two ideas -- more obvious in comparative graphics—that were essential to the representation of continuous distributions. One was the notion of a cumulative distribution as the graph of change in discrete quantities in an ordered sequence, the other was the concept of a continuous curve as the limit of ordered categories represented in the area under the curve.

A breakthrough in the first area was made by J. B.J. Fourier in 1821 (11). Fourier began with a bar graph representing the population of Paris by age groupings, then placed the bars one atop the other to form the ordinate of a line graph of a cumulative frequency distribution (given the name "ogive" by Galton in 1875).

A breakthrough in the other conceptual area came with the development of the histogram. In 1818, the German astronomer Bessel published a graphic table which employed numerals as in a histogram (9). The French mathematician Guerry applied Playfair's bar chart idea directly to continuous variables like age and time, for which he had data in ordered categories, and published histograms suggestive of various theoretical curves (14).

By mid-19th century, graphics had become an accepted part of the statistical discipline. The Third International Statistical Congress, meeting in Vienna in 1857, debated various graphic methods (10), and also organized the first exhibition display of graphs and cartograms. Statistical diagrams were introduced in school textbooks by the Frenchman Levasseur in 1868 (21). In 1872, the U.S. Congress appropriated the first money for a graphical treatment of statistical data, the cartograms of Ninth Census data which appeared in Statistics of Wealth and Public Indebtedness (26). Francis A. Walker, Superintendent of the Census, introduced two graphical forms—the age pyramid and the bilateral frequency polygon—in the Statistical Atlas published in 1874 (28).

Multivariate Distribution and Correlation

The rapid development of statistics in the latter half of the 19th century generated a new graphical problem: the representation of multivariate distributions and correlations. This problem was particularly crucial to vital statistics, which had need to treat interrelationships among at least three variables: population, age and time.

Léon Lalanne, French engineer and pioneer in mechanical computation, was first to reconstruct a three-way table as a two-dimensional contour map (19), published in 1845. Inspiration for Lalanne's idea came from nomographic tables, then essentially multiplication tables in contour form, which the French manufacturer Louis Pouchet published in 1795 (7), and which had been adopted by the French Artillery.

In 1869, Zeuner published a system that represented demographic trends as surfaces with three coordinates (29). Using axonometric projection, any "slice" of this surface, including 45-degree slices representing the history of various cohorts, could be shown in two dimensions. Ten years later, Luigi Perozzo, cartographic engineer for the Italian State Department of Statistics, produced a colored relief drawing (named a "stereogram") of Zeuner's theoretical surface, but based on actual data from the Swedish censuses of 1750-1875.

Lalanne had speculated that his contours of tables might be applied to geographical distributions, and 30 years later, in 1874, Vauthier published a map of Paris with contour lines showing densities of population (27). Galton cited the work of both Lalanne and Vauthier as inspirations for his normal correlation surface, published in 1886 (12). Two years later, Galton determined a coefficient of correlation by graphic means (13). The three-dimensional surfaces of Galton and Perozzo became favorites of instructors in probability theory and vital statistics, respectively, and—in various constructions of pasteboard and plaster of Paris—were standard equipment in statistical laboratories well into the 20th century.

Modern Developments

By the middle of the 20th century, interest in social and applied statistics in general, and graphical methods in particular, appeared to wane somewhat. Academicians increasingly turned their energies to theoretical concerns. This trend did not begin to reverse again until the mid-1960's, when developments in computer technology made possible the manipulation and analysis of large, multivariate data sets. Accompanying the renewed interest in data analysis was the development of computer graphics hardware and software, and advances in high-speed printing, xerography, microfilming, etc.

In the past 15 years, a spate of new quantitative graphical forms, which begin to exploit the modern technologies, have appeared in the technical literature of a variety of disciplines. As examples, one can cite Anderson's circular glyphs (30), the triangles of Pickett and White (36), Bachi's "graphical rational patterns" (32), the irregular polygons devised by Siegal and his collaborators (37), Andrews' Fourier form for generating multivariate plots (31), Chernoff's cartoons of human faces to represent multivariate data (34), and the color-coded bivariate matrix developed by the U.S. Bureau of the Census (33). John Tukey's various innovations for exploratory data analysis (39) seem particularly appropriate for the interactive capabilities of computer graphics, and have already been adapted for that purpose in various routines developed by the National Bureau of Economic Research.

The future of statistical graphics might be expected to lie in solutions to those problems generated or made tractable by computer and associated technologies. These problems include:

NINETEENTH CENTURY DEVELOPMENTS IN STATISTICAL GRAPHICS

1800 Use of coordinate paper in published research (graph of barometric variations) —Philosophical Magazine, Vol. 7, p. 357

Idea for continuous log of automatically recorded time-series graphs (temperature and barometric pressure) --A. Keith (English)

1801 Pie chart, circle graph--Wm. Playfair (English), Statistical Breviary

1811 Subdivided bar graph --F. H. Alexander von Humboldt (German)

1818 Graphic table, employing numerals as in a histogram -- Friedrich Wilhelm Bessel(German astronomer)

1819 Cartogram (map with shadings from black to white, showing distribution and intensity of illiteracy in France)--Charles Dupin (French geometer, statistician)

1820's An ever increasing number of scholarly publications begin to contain graphs which describe (but do not analyze) natural phenomena like magnetic variation, weather and tides

1821 Ogive or cumulative frequency curve (inhabitants of Paris by age groupings for 1817)--J.B.J. Fourier (French)

1828 Mortality curves from empirical data (for Belgium and France) --A. Quetelet (Belgian statistician)

1830-1835 Graphical analysis of natural phenomena begins to appear on a regular basis in scholarly publications, particularly in England

1831 Graph of frequency distribution--Quetelet

1832 Curve-fitting to scatter-plot; advocacy of graph paper as standard tool of science --J.F.W. Herschel (English)

1833 Histogram (crimes by age groupings, and by months) --A.M. Guerry (French)

1833 Rank lists , with lines showing shifts in rank order between categories (shifts in rank of crimes from one age group to the next)--A.M. Guerry

1838 Published graph of normal curve--Augustus De Morgan (English probabilist)

1841 Graph in Journal of London Statistical Society (established 1837)

1843 Logarithmic grid -- Léon Lalanne (French engineer)

Polar coordinates (frequency of wind directions)--Lalanne

Contour map of table(temperature x hour x month) --Lalanne

1846 Results of urn schemata as symmetrical histogram, with limiting normal curve--Quetelet

1847 Statistical map (tone wash) in Journal of London Statistical Society

1851 Map incorporating statistical diagrams (circles proportional to coal production) --Charles Joseph Minard (French)

1852 Graphics used in lawsuit (Germany)

1857 Discussion of graphical methods before International Statistical Congress --Third, in Vienna

Exhibition display of graphs and cartograms -- Third International Statistical Congress, Vienna

Polar area charts, known as "coxcombs" --Florence Nightingale (English), in anonymous publication for campaign to improve sanitary conditions of army

1863 Semi - logarithmic grid (percentage changes in commodities) -- S. Jevons (English)

1868 Three-dimensional population surface or "stereogram," with axonometric projection to show curves of various "slices"--Gustave Zeuner (German)

1868 Statistical diagrams in a school textbook (geography) -- Emile Levasseur (French), La France, avec ses Colonies

1872 U.S. Congressional appropriation for graphical treatment of statistics

Use of statistical graphics by U.S. government in census reports (cartograms of Ninth Census data)--Statistics of Wealth and Public Indebtedness

Classification of statistical graphic treatments by form--H. Schwabe(Ger.)

1874 Age pyramid (bilateral histogram) , bilateral frequency polygon --Francis A. Walker (Superintendent of U.S. Census), Statistical Atlas of U.S. Based on Results of Ninth Census

Population contour map (orthographic projection, of Paris) --L.L. Vauthier (French), cited by Galton as inspiration for normal correlation surface

1879 Stereogram (three-dimensional population pyramid) modeled on actual data (Swedish census , 1750-1875)--Luigi Perozzo (Italian)

Published instructions on how to use graph paper -- S. Jevons (Eng.), Principles of Science, 3rd Ed.

1883 Patent issued on logarithmic paper --England

1884 Pictogram--Michael George Mulhall (English) , Dictionary of Statistics

1885 Normal correlation surface--Galton, "Regression towards Mediocrity in Hereditary Structure"

1888 Correlation "coefficient" by graphic means--Galton

1895 Word "histogram" -- Karl Pearson (English) , lectures on graphical representation in statistics

1899 Idea for "log-square" paper ruled so normal probability curve appears as a straight line--Galton

(1) representation of large, multivariate data sets in two dimensions (as in Chernoff's "faces"), which has an informal history in medical diagnosis (37);

(2) representation of sampling and measurement error in traditional forms like bar charts, line graphs and statistical maps (for a discussion, see 35);

(3) representation of variables simultaneously with geographical and population distributions (what Tukey has called the "patch map" problem), as in Bachi's applications (32);

(4) representation of two or more intervally-measured variables in a single map, as in the color-coded cross-classification maps of the U.S. Census Bureau's Urban Atlas series (40); and

(5) more general development of graphics for use in computer-assisted analysis of large, multivariate data sets, as in Tukey's "exploratory data analysis."

In each of these areas, computer graphics applications have just begun to break away from imitations of forms--bar charts and line graphs, histograms and scatterplots -- which were commonplace a hundred years ago, and which do not seem inherently suited to dynamic or interactive applications. The potential uses of dynamic, three-dimensional and color graphics in quantitative applications remain largely unexplored. If the centuries-old struggle to use graphical representation in science and technology has any bearing on recent and future developments, then we can expect work in at least some of the problem areas to give way to more abstract and radically different forms--and perhaps even to new paradigmatic solutions.

REFERENCES

I. Early Graphics (Pre-1800)

1. Crome, August Friedrich Wilhelm. 1782. Producten-Karte von Europa.
2. Descartes, René. 1637. "La géométrie." Appendix to Discours de la Methode.
3. Halley, Edmund. 1686. "On the Height of Mercury in the Barometer at Different Elevations." Philosophical Transactions 16: 104-115.
4. _____. 1693. "An Estimate of Degrees of Mortality of Mankind, Drawn from Curious Tables of Births and Funerals at the City of Breslaw." Philosophical Transactions 17: 596-610.
5. _____. 1701. Description and Uses of a New and Correct SEA-CHART of the Whole World, Shewing Variations of the COMPASS.
6. Playfair, William. 1786. Commercial and Political Atlas.
7. Pouchet, Louis Ezéchiel. 1795. "Arithmétique Linéaire." Appendix to Echelle graphique des nouveaux poids, mesures et monnaies de la République française.
8. Priestley, Joseph. 1765. A Chart of Biography. (final edition, 1820)

II. Nineteenth Century

9. Bessel, Friedrich Wilhelm. 1818. Fundamenta Astronomiae pro anno MDCCLV.
10. Congrès International de Statistique. 1858. "Emploi de la cartographie et de la méthode graphique en général pour les besoins spéciaux de la statistique." Compte rendu de la troisième session: 192-97.
11. Fourier, J.B.J. 1821. "Notions generales sur la population." Recherches statistiques sur la Ville de Paris et le Department de la Seine I: lxvii.
12. Galton, Francis. 1885-86. "Regression towards Mediocrity in Hereditary Stature." Journal of the Anthropological Institute 15: 246-263.
13. _____. 1888. "Co-relations and Their Measurement, Chiefly from Anthropomorphic Data." Royal Society of London Proceedings 45: 135-145.
14. Guerry, André Michel. 1833. Essai sur la statistique morale de la France.
15. Herschel, William. 1800. "Experiments on Solar and Terrestrial Rays that Occasion Heat." Philosophical Transactions: 437-538.
16. Humboldt, F.H. Alexander von. 1811. Essai politique sur le royaume de la Nouvelle-Espagne.
17. Jevons, Stanley. 1863. A Serious Fall in the Value of Gold Ascertained, and its Social Effects Set Forth, with Two Diagrams.
18. _____. 1879. Principles of Science, 3rd ed.
19. Lalanne, Léon. 1845. "Sur la représentation graphique des tableaux météorologiques et des lois naturelles en général." Cours Complet de Météorologie de L.F. Kaemtz.
20. _____. 1846. Mémoire sur les tables graphiques et sur la géométrie anamorphique appliquées à diverses, questions qui se rattachent à l'art de l'ingénieur.
21. Levasseur, Pierre Émile. 1868. La France, avec ses Colonies. Contenant 174 cartes, coupes et figures explicatives insérées dans le texte.
22. Minard, Charles Joseph. 1861. Des Tableaux graphiques et des cartes figuratives.
23. Perozzo, Luigi. 1880. "Della rappresentazione grafica di una collettività di individui nella successione del tempo, e in particolare dei diagrammi a tre coordinate." Annali di Statistica 12: 1-16.
24. Philosophical Magazine. 1800. Vol. 7: 357.
25. Playfair, William. 1801. Statistical Breviary.
26. U.S. Bureau of the Census. 1872. Statistics of Wealth and Public Indebtedness.
27. Vauthier, L.L.1874. "Note sur une carte statistique figurant la répartition de la population de Paris." Comptes Rendus 78: 264-67.
28. Walker, Francis A. 1874. Statistical Atlas of the U.S. Based on Results of the Ninth Census.
29. Zeuner, Gustav. 1869. Abhandlungen aus der mathematischen Statistik.

III. Twentieth Century

30. Anderson, Edgar. 1960. "A Semigraphical Method for the Analysis of Complex Problems." Technometrics 2(3): 387-91.
31. Andrews, D.F. 1972. "Plots of High Dimensional Data." Biometrics 28: 125-36.
32. Bachi, Roberto. 1968. Graphical Rational Patterns. A New Approach to Graphical Presentation of Statistics.
33. Barabba, Vincent P.1975. "The Graphic Presentation of Statistics." U.S. Bureau of the Census, The Census Bureau, A Numerator and Denominator for Measuring Change: 89-103.
34. Chernoff, Herman. 1973. "The Use of Faces to Represent Points in k - Dimensional Space Graphically." Journal of the American Statistical Association 68(342): 361-68.
35. Fienberg, S.E., and L.A. Goodman. 1974. "Statistical Considerations." R.A. Van Dusen,ed., Social Indicators 1973: A Review Symposium.
36. Pickett, R., and B.W. White. 1966. "Constructing Data Pictures." Proceedings of the 7th National Symposium for Information Display: 75-81.

TWENTIETH CENTURY DEVELOPMENTS IN STATISTICAL GRAPHICS

1905 Lorenz curve (cumulative distribution by rank order, to facilitate study of concentrations) --M.O. Lorenz (U.S.)

c. 1910 Statistical diagrams in U.S. textbooks (graphs of temperature, population, in texts of arithmetic, algebra)

1913 Arithmetic probability paper, ruled so ogive appears as straight line (applied to problems of surface drainage) --Allen Hazen (U.S. engineer)

1913-1914 College course in statistical graphic methods--M. F.P. Costelloe (Dept. of Agricultural Engineering, Iowa State College)

1914 Published standards of graphic presentation for U.S.--American Society of Mechanical Engineers

Pictograms, uniform size (to replace bar graphs, pictograms of varying size)--W.C. Brinton(U.S.)

1916 Correspondence course in graphic methods (20 lessons, 50 dollars) --Frank J. Warne (U.S.)

1917 Published exposition of use of semi-logarithmic paper--James Field (U.S.)

1918-1933 Annual college course in statistical graphic methods--E.P. Cubberly (Stanford University)

1924 Social statistics graphics museum -- Social and Economic Museum, Vienna (Otto Neurath, Director)

1931 "Log square" paper --F.C. Martin and D.H. Leavens

1933 Standard statistical symbols established by government decree -- Soviet Union (for schools, public posters)

mid-1950's Application of cathode ray tube graphic terminals -- SAGE Air Defense System (U.S.)

1960 Circular glyphs , with rays to represent multivariate data -- Edgar Anderson (U.S.)

1962 Use of cathode ray tube graphic terminals in non-military environments -- Ivan E. Sutherland (U.S.) "Sketchpad," Lincoln Lab

1965 Improvements on histogram in analysis of counts, tail values--John Tukey

1966 Triangles to represent simultaneously four variables, using sides and orientation -- R. Pickett and B.W. White (U.S.)

1968 Systematic, tested patterns for graphic presentation--R. Bachi (Israel)

1969 Graphic innovations for exploratory data analysis (stem-and-leaf, box-and-whisker plots , hanging and suspended rootograms) --John Tukey

1971 Irregular polygon to represent multivariate data --J.H. Siegel, R.M. Goldwyn and H.P. Friedman

1972 Form of Fourier series to generate plots of multivariate data--D.F.Andrews

1973 Cartoons of human face to represent multivariate data--Herman Chernoff

U.S. Government chartbook devoted exclusively to social indicator statistics -- Social Indicators 1973, OMB

1974 Color-coded bivariate matrix to represent two intervally - measured variables in single map--U.S. Bureau of Census , Urban Atlas series

Weekly chartbook to brief U.S. President,Vice President on economic, social matters--Bureau of Census and OMB , at request of V.P. Nelson Rockefeller

37. Siegel, J.H., R.M. Goldwyn and H.P. Friedman. 1971. " Pattern and Process of the Evolution of Human Septic Shock." Surgery 70(2):232-45.
38. Tukey, John W. 1965. "The Future of Processes of Data Analysis." U.S. Army Research Office Report 65-3: 691-729.
39. _____. 1977. Exploratory Data Analysis.
40. U.S. Bureau of the Census and Manpower Administration. 1974. Urban Atlas, Series GE-80.
41. U.S. Office of Management and Budget. 1973. Social Indicators 1973.

IV. Secondary Sources

42. Beniger, James R.1975. "Annotated Bibliography: History of Quantitative and Statistical Graphics." Paper accompanying slide presentation, Annual Meeting of the American Statistical Association, Atlanta.
43. _____.1976. "From Stylus to Light Pen: Technology and Innovation in the Development of Statistical Graphics." Computers and Art 1(4).
44. _____.1976. "Science's 'Unwritten' History: The Development of Quantitative and Statistical Graphics."Paper presented to the Annual Meeting of the American Sociological Association. New York.
45. Funkhouser, H. Gray. 1936. "A Note on a Tenth-Century Graph." Osiris 1: 260-62.
46. _____. 1937. "Historical Development of the Graphical Representation of Statistical Data." Osiris 3(1): 269-405.
47. Hoff, H.E., and L.A. Geddes. 1959. "Graphic Recording Before Carl Ludwig:An Historical Summary." Archives Internationales d'Histoire des Sciences 12: 1-25.
48. Marey, E.J. 1878. La méthode graphique dans les sciences expérimentales et principalement en médecine.
49. Royston, Erica. 1956. "A Note on the History of Graphical Presentation of Data." Biometrika 43(3&4): 241-247.
50. Tilling , Laura. 1975. "Early Experimental Graphs." British Journal for the History of Science and Technology 8(30): 193-213.

TRIGONOMETRIC CONVERSION OF FALLIBLE DISTANCES INTO COORDINATES IN MDS

Purnell H. Benson, Rutgers Graduate School of Business Administration

In recent years, the problem of determining coordinates of points when interpoint distances are given has claimed attention of statisticians in the field of psychological measurement. The problem arises from a practical standpoint because human subjects are frequently able to report similarities between stimuli when they are unable to describe characteristics of the stimuli. From judgments of proximities between stimuli are calculated the coordinates of stimulus points. These coordinates, after suitable rotation and translation of axes, are in effect measurements of characteristics of the stimuli. The multidimensional distance scaling procedures of Torgerson (1952, 1958) and of Shepard (1962a, 1962b) and Kruskal (1964a, 1964b) have sought solutions to systems of quadratic equations with many sets of roots.

The problem of ascertaining the globally correct solution has not been resolved to the satisfaction of all concerned. Trying many different solutions to find the one with minimum stress upon the inputted distances is one answer. Another is to use the principal components procedure of Torgerson. Shepard (1974) expresses reservations about the adequacy of any existing procedures to furnish the globally correct solution except by repeated trial and error to find that set of coordinates which most closely fit the interpoint distances.

The present author has proposed a trigonometric solution for obtaining point coordinates from exact interpoint distances (1976b). The formulas for calculating coordinates of points from inputted interpoint distances are given in Table 1 for 4 points. The pairs of numbers in parentheses indicate interpoint distances. The triads of numbers designate angles. The hinge of the angle is an italicised number. For dihedral angles, the hinge involves more than one point.

Generalizing the formulas in Table 1 to larger numbers of points in more than 3 dimensions, letting $C_{M,N}$ be the $\underline{M}$th coordinate of the $\underline{N}$th point, $M<(N-1)$, we have

$$C_{M,N} = (1N)\ \sin(2\ \underline{1}\ N)\ \sin(3\ \underline{12}\ N)$$
$$\sin(4\ \underline{123}\ N)\ldots\cos(M\ \underline{123..M-1}\ N).$$

If $M = N-1$ for the last coordinate of the $\underline{N}$th point in the simplex, the terminal cosine term in the product is replaced by the sine of the same dihedral angle.

Also we have, with $L<M<N$

$$\cos(M\ \underline{123..L}\ N) = [\cos(M\ \underline{123..L-1}\ N) - \cos(L\ \underline{123..L-1}\ M)\ \cos(L\ \underline{123..L-1}\ N)]$$
$$/[\sin(L\ \underline{123..L-1}\ M)\ \sin(L\ \underline{123..L-1}\ N)]$$

These formulas have been tested for randomly selected coordinates of 20 points in 14 dimensions and have been found to be correct.

Application to Fallible Data

The formulas given can be directly applied to a network of error-free interpoint distances. The usual problem in scaling characteristics of stimuli is that the respondents give inaccurate proximity judgments which do not form an exact system. Depending upon which interpoint distances are used, different sets of coordinates result, if indeed any calculation at all is mathematically possible.

A plausible procedure is to obtain exact solutions to alternative sets of points and then to average the resulting coordinates transformed to the same axes of orientation. A machine program was prepared which takes simplexes of

Table 1

COORDINATES FOR 4 POINTS IN 3 DIMENSIONS

	Coordinate		
Point	1	2	3
1	0	0	0
2	(12)	0	0
3	$(13)\cos(2\underline{1}3)$	$(13)\sin(2\underline{1}3)$	0
4	$(14)\cos(2\underline{1}4)$	$(14)\sin(2\underline{1}4)\cos(3\underline{12}4)$	$(14)\sin(214)\sin(3\underline{12}4)$

points in all possible sets to calculate average coordinates from the different possible exact solutions. The program, TRIVCOR, is available.

First is selected the largest simplex which can be formed from the points in the number of dimensions for which a solution is pursued. Then each possible simplex in that number of dimensions has its coordinates calculated and referred to the largest simplex for the axes on which the averaging of coordinates is done.

Coordinates between -1.0 and 1.0 were randomly chosen for 8 points in 1, 2, 3, 4, and 5 dimensions. The distances between the points were calculated and then degraded by adding error quantities whose absolute averages are respectively .1, .2, .4 and .8, in separate computations. The 4 levels of error for 5 different levels of dimensions were pursued in 20 different computations to fit coordinates to simulated data. The coordinates obtained were then reconverted to interpoint distances whose errors are then compared with the original distances before their degradation by random errors.

In some cases, the error degradation of the interpoint distances led to impossible configurations, such as one side of a triangle being greater than the sum of the other two sides. An adjustment computation was programmed and performed which modified the system of interpoint distances by small increments until a consistent set of distances was obtained. Starting with the largest interpoint distance, each set of 3 was tested for consistency with a tolerance of .01. If the tolerance condition was not met, the longer side was reduced by .005 and the shorter sides were increased by .0025 each. The consistency adjustment was repeated iteratively until a consistent system of all of the interpoint distances was obtained. If the distances are not consistent, the trigonometric calculation gives defective results.

This consistent (but not exact) set of interpoint distances then formed the starting point for the calculation and averaging of sets of coordinates for all possible simplexes. With 8 points, the number of simplexes is 28 for 1 dimension, 56 for 2 dimensions, 70 for 3 dimensions, 56 for 4 dimensions, and 28 for 5 dimensions. These are the numbers of separate calculations of coordinates made before averaging coordinates for the different numbers of dimensions.

The results from the 20 different computations are given in Table 2. The coordinates for interpoint distances with small random errors, of the order of 10 per cent, can be satisfactorily estimated for points in one or two dimensions. For larger errors or larger numbers of dimensions, the estimation of coordinates becomes unsatisfactory.

Table 2

AVERAGE ERRORS IN OUTPUTTED INTERPOINT DISTANCES RECOVERED FROM COORDINATES CALCULATED FROM DISTANCES BETWEEN RANDOM POINTS

Number of dimensions	Average error introduced in distances at random	Average original distance	Average error in inputted distance	Average error in outputted distance recovered
1	.10	.79	.09	.09
	.20	.52	.19	.13
	.40	.69	.35	.24
	.80	.71	.63	.58
2	.10	1.16	.10	.12
	.20	1.08	.19	.20
	.40	1.09	.35	.23
	.80	1.16	.72	.32
3	.10	1.41	.10	.66
	.20	1.27	.15	.12
	.40	1.06	.36	.56
	.80	1.44	.78	.48
4	.10	1.49	.10	.26
	.20	1.86	.20	1.38
	.40	1.65	.50	1.67
	.80	1.43	.65	.62
5	.10	1.81	.10	.43
	.20	1.88	.23	.89
	.40	2.04	.40	1.22
	.80	1.96	.75	1.64

Possibly the results from averaging coordinates of alternative simplexes can form the starting solution for the iterative solution of the quadratic system of equations. The most dependable procedure appears to be to use the method of descent described by the author (1976a). In this method of descent, the exact calculation of coordinates from alternative simplexes may be expected to improve the computational efficiency.

REFERENCES

Benson, P. H. Using simplex relationships in multidimensional scaling. Working paper, presented at the program of the Psychometric Society and the Mathematical Psychology Group April 2, 1976. (a)

Benson, P. H. A direct procedure for calculating point coordinates from interpoint distances. Publication pending, expected 1976 or 1977. (b)

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27. (a)

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 28-42. (b)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-140. (a)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962, 27, 219-246. (b)

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. Psychometrika, 1974, 39, 373-421.

Torgerson, W. S. Multidimensional scaling: I. theory and method. Psychometrika, 1952, 17, 401-419.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

ON SOME NONPARAMETRIC TESTS FOR PROFILE ANALYSIS OF SEVERAL MULTIVARIATE SAMPLES

V. P. Bhapkar, University of Kentucky
Kenneth W. Patterson, Ethicon

Summary

For the profile analysis of several multivariate samples in a nonparametric framework without assuming normality, a certain appropriate hypothesis of parallelism of population profiles is formulated. A class of test criteria is obtained to test such an hypothesis. The overall χ^2-statistic arising from differences among populations is partitioned into two components--the first due to the "interaction" between populations and variables, and the remainder due to the "pure main effects" from the populations. Some theoretical properties of the criteria are established and, finally, simulation studies are carried out to investigate the performance of these criteria for small or moderate size samples.

Introduction

Let $\underset{\sim}{X}_{ij}' = (X_{ij}^{(1)}, X_{ij}^{(2)}, \ldots, X_{ij}^{(p)})$, $j=1,2,\ldots,n_i$ be independent random vectors from the i-th population with nonsingular continuous c.d.f. F_i. Assume that we have such independent samples from k populations for $i=1,2,\ldots,k$ with a total sample of size $N = \Sigma_i n_i$ on p variables. In the parametric framework it is usually assumed that the i-th population is p-variate normal with mean $\underset{\sim}{\mu}' = (\mu_i^{(1)},\ldots,\mu_i^{(p)})$ and common nonsingular covariance matrix $\underset{\sim}{\Sigma}$. The hypothesis of homogeneity is then

$$H_0: \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = \ldots = \underset{\sim}{\mu}_k,$$

while the hypothesis of parallelism of population profiles is usually formulated as

$$H_1: \mu_i^{(1)}-\mu_1^{(1)} = \ldots = \mu_i^{(p)}-\mu_1^{(p)}. \quad i=2,\ldots,k.$$

The formulation H_1 arises naturally by considering population profiles obtained by plotting means $\mu_i^{(\alpha)}$ against variable $\alpha=1,\ldots,p$ for each i.

Fig. 1: Population Profiles in terms of Means

The corresponding sample profiles are obtained on replacing population means by sample means. The appropriate test criteria for H_0 and H_1 respectively are available in standard texts (see, e.g., [5]).

It is desirable to construct suitable nonparametric analogs in order to discard the stringent assumption of p-variate normality, especially in situations where we can observe ordinal data without precise numerical measurements X's. Nonparametric tests are already available in the statistical literature (see, e.g., [2], [6], [7]) for the homogeneity hypothesis

$$H_0^*: \ F_1=F_2=\cdots=F_k,$$

which is the obvious analog of H_0 in the nonparametric case. In this paper a suitable nonparametric analog of H_1, viz. H_1^*, is first formulated and, then, asymptotic chi-square criteria are offered to test H_1^*.

Preliminaries

The nonparametric tests of H_0^* presented independently by Bhapkar [2] and Suguira [7] are based on the technique of generalized U-statistics. Such tests were developed initially for the univariate case by Bhapkar ([1]).

In the multivariate case the generalized U-statistic $U_i^{(\alpha)}$ corresponding to a function $\phi_i^{(\alpha)}$ of k arguments is defined by

$$U_i^{(\alpha)} = \frac{1}{\prod_{j=1}^{k} n_j} \sum_{t_1=1}^{n_1} \cdots \sum_{t_k=1}^{n_k} \phi_i^{(\alpha)}(\underset{\sim}{X}_{1t_1},\ldots,\underset{\sim}{X}_{kt_k}), \quad (2)$$

$\alpha = 1,\ldots,p$ and $i = 1,\ldots,k$. Let

$$\underset{\sim}{U}_1' = (U_i^{(1)},\ldots,U_i^{(p)})$$

and

$$\underset{\approx}{U}' = (\underset{\sim}{U}_1',\ldots,\underset{\sim}{U}_k').$$

We assume that $\phi_i^{(\alpha)}$ is a specific rank function, say ϕ, comparing the α-th components of the i-th argument against the other k-1. Thus, we assume

$$\phi_i^{(\alpha)}(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_k) = \phi(r_i^{(\alpha)}), \quad (3)$$

where $r_i^{(\alpha)}$ is the rank of $x_i^{(\alpha)}$ among $\{x_j^{(\alpha)}, j=1,\ldots,k\}$. In view of continuity assumption, with probability one there are no ties. Note that the functions considered by Bhapkar [2] and Suguira [7] are special cases of functions satisfying (3).

Let $\underset{\sim}{F}' = (F_1,\ldots,F_k)$ and define

$\eta^{(\alpha)}(\underset{\sim}{F})=E(U_i^{(\alpha)})=E\phi_i^{(\alpha)}(\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_k)$, where $\underset{\sim}{X}_i$'s represent independent random vectors with c.d.f. F_i's respectively. Then we have

$$\eta_i^{(\alpha)}(\underset{\sim}{F}) = \sum_{j=1}^{k} \phi(j)P[R_i^{(\alpha)} = j] = \sum_{j=1}^{k} \phi(j)\nu_{ij}^{(\alpha)}(\underset{\sim}{F}); \quad (4)$$

here $R_i^{(\alpha)}$ is the rank of $X_i^{(\alpha)}$ among $\{X_j^{(\alpha)}, j=1,\ldots,k\}$ and

$$\nu_{ij}^{(\alpha)}(\underset{\sim}{F}) = P[R_i^{(\alpha)} = j], \quad (5)$$

with the probabilities computed under $\underset{\sim}{F}$.

Suppressing $\underset{\sim}{F}$ for the moment, let

$$\underset{\sim}{\eta}_i' = (\eta_i^{(1)},\ldots,\eta_i^{(p)}),\ \underset{\sim}{\eta}' = (\underset{\sim}{\eta}_1',\ldots,\underset{\sim}{\eta}_k').$$

Note that under H_0^*, $\nu_{ij}^{(\alpha)} = 1/k$ for all $\alpha = 1,\ldots,p$, and $i,j = 1,\ldots,k$, so that $\underset{\sim}{\eta}(\underset{\sim}{F}) = \phi\underset{\sim}{j}$, where $\underset{\sim}{j}$ is a column-vector of appropriate order with all elements 1 and

$$\phi = \frac{1}{k}\sum_{j=1}^{k}\phi(j). \quad (6)$$

Now it is known that if $n_i \to \infty$ in such a way that $n_i/N \to p_i$, where $N = \Sigma_i n_i$, $0 < p_i < 1$, $i=1,\ldots,k$, then

$$E(\underset{\sim}{U}_n) = \underset{\sim}{\eta}(\underset{\sim}{F}),\quad V(\underset{\sim}{U}_n) = \frac{1}{N}\underset{\sim}{T}(\underset{\sim}{F}) + O(N^{-3/2}) \quad (7)$$

and

$$N^{1/2}(\underset{\sim}{U}_n - \underset{\sim}{\eta}(\underset{\sim}{F})) \xrightarrow{\mathcal{L}} \mathcal{N}(\underset{\sim}{0}, \underset{\sim}{T}(\underset{\sim}{F})),$$

for any $\underset{\sim}{F}$. Here the subscript $\underset{\sim}{n}$ denotes the vector of sample sizes on which $\underset{\sim}{U}$ is based, V denotes the covariance matrix, $\mathcal{L}$ denotes convergence in distribution, $\mathcal{N}$ the normal vector of appropriate dimensions.

It was shown in [2] that under $H_0^*(1)$,

$$\underset{\sim}{\eta}(F) = \phi\underset{\sim}{j},\ \underset{\sim}{T}(F) = \underset{\sim}{\Sigma} \otimes \underset{\sim}{\rho}(F), \quad (8)$$

where $\underset{\sim}{A}\otimes\underset{\sim}{B} = [a_{ij}\underset{\sim}{B}]$, and $\underset{\sim}{\Sigma} = [\sigma_{ij}]$ is given by

$$\underset{\sim}{\Sigma} = \frac{\mu}{(k-1)^2}\{q\underset{\sim}{J} + k^2\underset{\sim}{\Delta} - k\underset{\sim}{q}\underset{\sim}{j}' - k\underset{\sim}{j}\underset{\sim}{q}'\} \quad (9)$$

with $\underset{\sim}{J} = [1]$, $\underset{\sim}{\Delta}$ diagonal $(p_i^{-1}, i=1,\ldots,k)$, $q = \Sigma_i p_i^{-1}$ and $\underset{\sim}{q}' = (p_1^{-1},\ldots,p_k^{-1})$. Also $\underset{\sim}{\rho}$ is a matrix of correlation coefficients $\rho_{\alpha\beta}$ between $\phi_i^{(\alpha)}(\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_k)$ and $\phi_i^{(\beta)}(\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_k)$, where $\underset{\sim}{X}$'s and Y's are independent with common c.d.f. F except that $\underset{\sim}{X}_i = \underset{\sim}{Y}_i$, and

$$\mu = E[\psi^2(X_i^{(\alpha)})] - [E(\psi(X_i^{(\alpha)}))]^2 \quad (10)$$

where

$$\psi(x_i^{(\alpha)}) = E\{\phi_i^{(\alpha)}(\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_k) | \underset{\sim}{X}_i = \underset{\sim}{x}_i\}.$$

It can be shown that in view of condition (3) μ does not depend on the common F under H_0^*; however it does depend on the function ϕ. Explicit values of μ are given in [2] for some specific functions ϕ.

Also it has been shown in [2] that if, under H_0^*, the common F is nonsingular (in the sense that the whole probability mass is not contained in any lower-dimensional space) then $\underset{\sim}{\rho}(F)$ is nonsingular. Then the matrix $\hat{\underset{\sim}{\rho}}$ of consistent estimators is also nonsingular with probability approaching one as all $n_i \to \infty$.

It was shown in [2] that

$$T_0 = \frac{N(k-1)^2}{\mu k^2}\sum_{i=1}^{k} p_i(\underset{\sim}{U}_i - \bar{\underset{\sim}{U}})\hat{\underset{\sim}{\rho}}^{-1}(\underset{\sim}{U}_i - \bar{\underset{\sim}{U}}), \quad (11)$$

where $p_i = n_i/N$ and $\bar{\underset{\sim}{U}} = \Sigma_i p_i \underset{\sim}{U}_i$, has a limiting $\chi^2(p(k-1))$ distribution under H_0^*. Explicit statistics denoted by V,B,L and W were offered as possible nonparametric test criteria (for the hypothesis H_0^*) corresponding to (i) $\phi_V(r)=1$, if r=1, and 0 otherwise, (ii) $\phi_B(r) = 1$, if r=k, and 0 otherwise, (iii) $\phi_L(1) = -1$, $\phi_L(k) = 1$ and $\phi_L(r) = 0$ otherwise, and (iv) $\phi_W(r) = r$.

Suguira [7] considered the class of functions

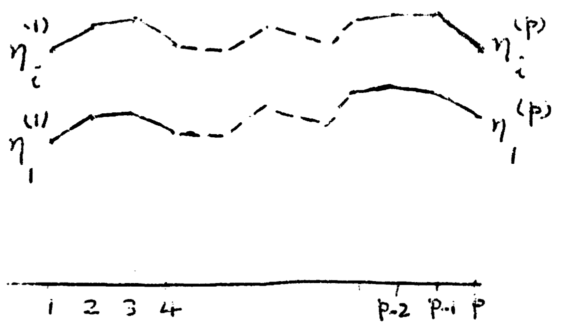
$$\phi_i^{(\alpha)}(x_1,\ldots,x_k) = \frac{(j-1)_r}{(k-1)_r} - \frac{(k-j)_s}{(k-1)_s}, \quad (12)$$

where j is the rank of $x_i^{(\alpha)}$ among $\{x_\ell^{(\alpha)}, \ell=1,\ldots,k\}$, and $(a)_r = a!/(a-r)!$. In view of (3), this function can be expressed as a member of the class $\{\phi_{r,s}, r,s = 0,1,\ldots,k-1\}$ with $\phi_{r,s}(j)$ denoting the right side of (12). The choices (i) (0,k-1), (ii) (k-1, 0), (iii) (k-1, k-1) and (iv) (1, 1) for (r, s), respectively, are essentially equivalent to ϕ_V, ϕ_B, ϕ_L and ϕ_W respectively. His statistic is essentially the same as (11) except that he uses somewhat different estimates for $\underset{\sim}{\rho}$. We may note here, however, that his estimates are consistent only under H_0^* while those in [2] are valid for any $\underset{\sim}{F}$ and hence the latter are to be preferred.

Nonparametric Parallelism Hypothesis

First we want to formulate an appropriate nonparametric analogue of the hypothesis H_1 of parallelism of profiles. In the parametric case the profiles are defined in terms of population means as in Figure 1. In the more general nonparametric case we can similarly define the population profiles in terms of quantities $\eta_i^{(\alpha)}$ which are the expected values of rank functions ϕ in (3). For each i, $\eta_i^{(\alpha)}$ indicates the relative rank location (for the specific ϕ function used) of the i-th population among the k populations with respect to the component α. Plotting these for various α would give the profile of the i-th population.

Fig. 2: Population Profiles in terms of η's



The corresponding sample profiles can similarly be plotted in terms of statistics $U_i^{(\alpha)}$, $\alpha=1,\dots,p$ for each $i=1,\dots,k$.

The obvious way to define the parallelism hypothesis is to require

$$\eta_i^{(1)} = \eta_i^{(2)} = \cdots = \eta_i^{(p)}$$

for each i. However, we note from (4) that η's do depend on the specific function ϕ. It is not desirable to formulate the hypothesis itself oriented towards a particular function ϕ. Rather we would prefer a formulation that works no matter which ϕ function is used. With this point in view, we now give the following definition:

Definition. The populations $F_1, F_2, \dots, F_k$ are said to have parallel profiles if $\underset{\sim}{F} = (F_1,\dots F_k)$ satisfies

$$H_1^*: \ \nu_{ij}^{(1)}(\underset{\sim}{F}) = \cdots = \nu_{ij}^{(p)}(\underset{\sim}{F}), \ i,j = 1,\dots,k, \tag{13}$$

where $\nu_{ij}^{(\alpha)}(\underset{\sim}{F})$ is defined by (5).

One might wonder whether H_1 and H_1^* are equivalent in some sense under the normality assumption. The answer is no except possibly the special case where the variances $\sigma_{\alpha\alpha}$ of $X_i^{(\alpha)}$ are the same for all $\alpha=1,\dots,p$. We prove here only the weaker statement:

Lemma. If $\underset{\sim}{X}_1,\dots,\underset{\sim}{X}_k$ are independent $\mathcal{N}(\underset{\sim}{\mu}_i, \underset{\sim}{\Sigma})$, respectively, and the diagonal elements of $\underset{\sim}{\Sigma}$ are equal, then H_1 implies H_1^*.

Proof: Note that

$$\nu_{ij}^{(\alpha)}(\underset{\sim}{F}) = P[R_i^{(\alpha)}=j] = \sum_c P[\text{Each of } \{X_{i\ell}^{(\alpha)}, \ \ell=1,\dots,j-1\} < X_i^{(\alpha)} < \text{Each of } \{X_{im}^{(\alpha)}, \ m=j+1,\dots,k\}]$$

$$= \sum_c P[\text{Each of } \{Y_{i\ell}^{(\alpha)}+\mu_{i\ell}^{(\alpha)}-\mu_i^{(\alpha)}\} < Y_i^{(\alpha)} < \text{Each of } \{Y_{im}^{(\alpha)} + \mu_{im}^{(\alpha)}-\mu_i^{(\alpha)}\}]; \tag{14}$$

here $\sum_c$ denote the sum over $\binom{k-1}{j-1}$ combinations of subscripts $i\ell$, $\ell = 1,\dots,j-1$ chosen out of k-1 distinct subscripts $i\ell$, $\ell=1,\dots,k$ (except j) (denoting integers $1,\dots,k$ except i).

Now $Y_i^{(\alpha)}$ for $\alpha = 1,\dots,p$ and $i = 1,\dots,k$ are independent and identical normal variables. If the condition H_1 is satisfied, we see from (14) that $\nu_{ij}^{(\alpha)}(\underset{\sim}{F})$ does not depend on α and hence, then, H_1^* is satisfied.

In fact normality as such is not used at all except for the fact that $\underset{\sim}{\mu}_i$ are location parameters. By using essentially the same argument we have thus proved the

Theorem 1. Suppose $\underset{\sim}{X}_1,\dots,\underset{\sim}{X}_k$ are independent with c.d.f.

$$F_i(\underset{\sim}{x}) = F(\underset{\sim}{x}-\underset{\sim}{\mu}_i), \quad i=1,\dots,k \tag{15}$$

for some continuous F, and assume that the marginal c.d.f.'s $F^{(\alpha)}$, $\alpha=1,\dots,p$, of F are identical, then the conditon H_1 implies the condition H_1^*.

Test of H_1^*

In order to test H_1^* we now propose the statistic

$$T_1 = \frac{N(k-1)^2}{\mu k^2} \sum_{i=1}^{k} p_i(\underset{\sim}{U}_i-\underset{\sim}{\bar{U}})'[\hat{\underset{\sim}{\rho}}^{-1}-\hat{\gamma}\hat{\underset{\sim}{\rho}}^{-1}\underset{\sim}{J}\hat{\underset{\sim}{\rho}}^{-1}](\underset{\sim}{U}_i-\underset{\sim}{\bar{U}})$$

$$= T_0 - T_2, \tag{16}$$

where T_0 is the statistic (11) for H_0^*,

$$T_2 = \frac{N(k-1)\hat{\gamma}}{\mu k^2} \sum_{i=1}^{k} p_i(\underset{\sim}{U}_i-\underset{\sim}{\bar{U}})'\hat{\underset{\sim}{\rho}}^{-1}\underset{\sim}{J}\hat{\underset{\sim}{\rho}}^{-1}(\underset{\sim}{U}_i-\underset{\sim}{\bar{U}}), \tag{17}$$

and $\hat{\gamma} = 1/\underset{\sim}{j}'\hat{\underset{\sim}{\rho}}^{-1}\underset{\sim}{j}$. T_1 is to be regarded as a large-sample $\chi^2((p-1)(k-1))$ criterion for H_1^* while T_2 as a $\chi^2(k-1)$ criterion for testing H_0^*, assuming H_1^*, i.e., for testing the 'pure' differences among the populations after eliminating from T_0 the interaction contribution, if any.

It may be noted here that if $\underset{\sim}{P}$ is any (p-1)xp matrix of rank p-1 satisfying $\underset{\sim}{P}\underset{\sim}{j} = \underset{\sim}{0}$, then

$$\underset{\sim}{\rho}^{-1} - \gamma\underset{\sim}{\rho}^{-1}\underset{\sim}{J}\underset{\sim}{\rho}^{-1} = \underset{\sim}{P}'(\underset{\sim}{P}\underset{\sim}{\rho}\underset{\sim}{P}')^{-1}\underset{\sim}{P},$$

where $\underset{\sim}{\rho}$ is a positive definite matrix and $\gamma = 1/\underset{\sim}{j}'\underset{\sim}{\rho}^{-1}\underset{\sim}{j}$. Since $\underset{\sim}{\rho}$ is a non-singular correlation matrix, it is positive definite, and so is $\hat{\underset{\sim}{\rho}}$ with probability tending to one as $n_i \to \infty$. Thus, we may also express T_1 as

$$T_1 = \frac{N(k-1)^2}{\mu k^2} \sum_{i=1}^{k} p_i(\underset{\sim}{U}_i-\bar{\underset{\sim}{U}})' \underset{\sim}{P}'(\underset{\sim}{P}\hat{\underset{\sim}{\rho}}\underset{\sim}{P}')^{-1}\underset{\sim}{P}(\underset{\sim}{U}_i-\bar{\underset{\sim}{U}}). \quad (18)$$

It is straightforward to show that, if H_0^* holds, $T_1 \xrightarrow{\mathcal{L}} \chi^2((p-1)(k-1))$ and $T_2 \xrightarrow{\mathcal{L}} \chi^2(k-1)$; this will also follow from Theorem 3 established in the next section. However, what we would like to have if possible is the stated limiting distribution of T_1 under H_1^* alone. This does not seem to be possible by the present approach (and perhaps by any other approach) without discarding the relative simple form of the statistic. Note in (7) that in general the limiting covariance matrix $\underset{\sim}{T}$ is a pk x pk matrix of functionals depending on $\underset{\sim}{F}$. It is only under H_0^* that $\underset{\sim}{T}$ had the structure $\underset{\sim}{\Sigma} \otimes \underset{\sim}{\rho}$, where $\underset{\sim}{\Sigma}$ is known, and now $\underset{\sim}{\rho}$ is a pxp matrix of functionals depending on common F. Discarding the Kornecker product structure would make it necessary to estimate all terms of $\underset{\sim}{T}$, thus making the computation much more involved. However, as we shall show shortly, the use of concept of 'local alternatives' to H_0^* still makes it possible to justify the use of statistic T_1 for testing H_1^*.

We now state here without proof a Theorem which establishes consistency of the T_0, T_1 and T_2 tests for appropriate alternatives. The reader is referred to [3] for further details.

Theorem 2. Let T_0, T_1 and T_2 be defined as (11), (16) and (17) for functions $\phi_i^{(\alpha)}$ satisfying (3). If $n_i \to \infty$ in such a way that $n_i/N \to p_i$, $0 < p_i < 1$, then

(i) $T_0 \xrightarrow{P} \infty$ iff $\underset{\sim}{F} \notin \{\underset{\sim}{F} | \Sigma_j \phi(j)\nu_{ij}^{(\alpha)}(\underset{\sim}{F})$ is independent of i and α, $i = 1,\ldots,k$, $\alpha = 1,\ldots,p\}$,

(ii) $T_1 \xrightarrow{P} \infty$ iff $\underset{\sim}{F} \notin \{\underset{\sim}{F} | \Sigma_j \phi(j)\nu_{ij}^{(\alpha)}(\underset{\sim}{F})$ is independent of α $= 1,\ldots,p$ for each $i = 1,\ldots,k\}$

and, if $\underset{\sim}{\rho}^{-1} = [\rho^{\alpha\beta}]$, then

(iii) $T_2 \xrightarrow{P} \infty$ iff $\underset{\sim}{F} \notin \{\underset{\sim}{F} | \Sigma_j \phi(j)\Sigma_{\alpha,\beta}\rho^{\alpha\beta}\nu_{ij}^{(\alpha)}(\underset{\sim}{F})$ is independent of i$\}$.

Remark. We thus note here that the tests T_0, T_1 designed for H_0^*, H_1^* respectively are consistent only against alternatives to the hypotheses 'effectively' being tested viz.

$$H_{0\phi}^*: \quad \sum_{j=1}^{k} \phi(j)\nu_{ij}^{(\alpha)}(\underset{\sim}{F}) \text{ is independent of i and } \alpha$$

and

$$H_{1\phi}^*: \quad \sum_{j=1}^{k} \phi(j)\nu_{ij}^{(\alpha)}(\underset{\sim}{F}) \text{ is independent of } \alpha,$$

depending on the function ϕ used for T's. Of course this undesirable feature of nonparametric tests is usually unavoidable, e.g., Mann-Whitney test, Sign test, Kruskal-Wallis test all suffer from a similar disadvantage.

Note also that if H_1^* is accepted, i.e. $\nu_{ij}^{(\alpha)}$ is independent of α, then $T_2 \xrightarrow{P} \infty$ unless $\Sigma_j \phi(j)\nu_{ij}(\underset{\sim}{F})$ is independent of i which is precisely the condition for $T_1 \not\xrightarrow{P} \infty$ assuming H_1^*.

Asymptotic Distributions

In the previous Theorem we have found the class of fixed alternatives $\underset{\sim}{F}' = (F_1,\ldots,F_k)$ for which the tests are consistent, i.e., for which the power of the respective test tends to 1 as $n_i \to \infty$. We shall now find the limiting distributions of T_0, T_1 and T_2 under the sequence of Pitman location alternatives

$$H_N: \quad F_{iN}(\underset{\sim}{x}) = F(\underset{\sim}{x} - N^{-1/2}\underset{\sim}{\delta}_i), \quad i=1,\ldots,k \quad (19)$$

where the $\underset{\sim}{\delta}_i$'s are not all equal, and $\Sigma_i \underset{\sim}{\delta}_i = \underset{\sim}{0}$. Let

$$\gamma_i^{(\alpha)}(F) = k\,\delta_i^{(\alpha)} q^{(\alpha)}(\phi, F), \quad \underset{\sim}{\gamma}_i' = (\gamma_i^{(1)},\ldots,\gamma_i^{(p)}),$$
$$\underset{\sim}{\gamma}' = (\gamma_1',\ldots,\gamma_k'), \quad (20)$$

where

$$q^{(\alpha)}(\phi,F) = \sum_{j=1}^{k} \phi(j)\left[\binom{k-2}{j-2} a^{(\alpha)}(j-2,k-j,F) - \binom{k-2}{j-1} a^{(\alpha)}(j-1,k-j-1,F)\right]$$

and

$$a^{(\alpha)}(b,c,F) = \int_{-\infty}^{\infty} [F^{(\alpha)}(y)]^b\, [1-F^{(\alpha)}(y)]^c\, f^{(\alpha)}(y)\, dF^{(\alpha)}(y).$$

The result concerning limiting distribution is now stated here without proof. The reader is referred to [3] for further details.

Theorem 3. Consider the sequence $\{H_N\}$ of distributions $\{\underset{\sim}{F}_N\}$ given by (9) and assume that $F^{(\alpha)}$ is differentiable and has a bounded derivative $f^{(\alpha)}$ almost everywhere, $\alpha = 1,\ldots,p$. Suppose further that there exist functions $g^{(\alpha)}$ such that for sufficiently small h

$$\left|\frac{F^{(\alpha)}(x+h) - F^{(\alpha)}(x)}{h}\right| \le g^{(\alpha)}(x)$$

for almost all x, and $\int_{-\infty}^{\infty} g^{(\alpha)}(x)dF^{(\alpha)}(x) < \infty$.

Then as $n_i \to \infty$, so that $n_i/N \to p_i$, $0 < p_i < 1$,

$$T_0 \xrightarrow{\mathcal{L}} \chi^2(p(k-1), \lambda_0(\phi,\underset{\sim}{\delta},F)),$$
$$T_1 \xrightarrow{\mathcal{L}} \chi^2((p-1)(k-1), \lambda_1(\phi,\underset{\sim}{\delta},F)) \qquad (21)$$
$$T_2 \xrightarrow{\mathcal{L}} \chi^2((k-1), \lambda_2(\phi,\underset{\sim}{\delta},F))$$

where

$$\lambda_0(\phi,\underset{\sim}{\delta},F) = \frac{(k-1)^2}{\mu k^2}\sum_{i=1}^{k} p_i(\underset{\sim}{\gamma}_i - \bar{\underset{\sim}{\gamma}})'\underset{\sim}{\rho}^{-1}(\underset{\sim}{\gamma}_i - \bar{\underset{\sim}{\gamma}}),$$

$$\lambda_1(\phi,\underset{\sim}{\delta},F) = \frac{(k-1)^2}{\mu k^2}\sum_{i=1}^{k} p_i(\underset{\sim}{\gamma}_i - \bar{\underset{\sim}{\gamma}})'[\underset{\sim}{\rho}^{-1} - \gamma\underset{\sim}{\rho}^{-1}\underset{\sim}{J}\underset{\sim}{\rho}^{-1}](\underset{\sim}{\gamma}_i - \bar{\underset{\sim}{\gamma}}),$$

and

$$\lambda_2(\phi,\underset{\sim}{\delta},F) = \lambda_0(\phi,\underset{\sim}{\delta},F) - \lambda_1(\phi,\underset{\sim}{\delta},F).$$

Now we are in a position to identify the sequences $\{H_N\}$ of distributions $\{\underset{\sim}{F}_N\}$ for which the criteria have limiting null distributions.

Theorem 4. Assume conditions of Theorem 3 and suppose that $q^{(\alpha)}(\phi,F) \neq 0$. Then

(i) $T_0 \xrightarrow{\mathcal{L}} \chi^2(p(k-1))$ iff H_0^* holds;

furthermore, if $F^{(\alpha)} = F^{(\beta)}$ for all $\alpha \neq \beta$, then

(ii) $T_1 \xrightarrow{\mathcal{L}} \chi^2((p-1)(k-1))$ iff $\delta_i^{(1)} = \cdots = \delta_i^{(p)}$, $i=1,\ldots,k$

and

(iii) $T_2 \xrightarrow{\mathcal{L}} \chi^2(k-1)$ iff H_0^* holds, assuming $\delta_i^{(\alpha)} = \delta_i^{(\beta)}$ for all $\alpha \neq \beta$.

Remark. Note here that T_1 has a limiting central χ^2-distribution under $\{H_N\}$ only with side conditions that the form of marginal distributions is the same for all components and the location parameters are in the same relative position for each component α for the i-th population. The latter condition is similar to the statement of H_1 in the parametric case. However, here we require in addition the equality of all marginal distributions except for location parameters. This conditon is similar (in fact, equivalent) to the condition of 'commensurability' required in the parametric case (see [5]) for profile analysis to be meaningful.

Finally, in this section, we present the form of q's for some ϕ-functions referred to in section 2:

(i) $\phi = \phi_V$, $q^{(\alpha)}(\phi_V,F) = -a^{(\alpha)}(0,k-2,F)$

(ii) $\phi = \phi_B$, $q^{(\alpha)}(\phi_B,F) = a^{(\alpha)}(k-2,0,F)$

(iii) $\phi = \phi_L$, $q^{(\alpha)}(\phi_L,F) = a^{(\alpha)}(0,k-2,F) + a^{(\alpha)}(k-2,0,F)$

and

(iv) $\phi = \phi_W$, $q^{(\alpha)}(\phi_W,F) = \int_{\infty}^{\infty} f^{(\alpha)}(y)dF^{(\alpha)}(y)$.

Concluding Remarks

We have thus established, first of all, consistency of the three tests T_0, T_1, and T_2 for a specific ϕ function against alternatives to H_0^*, H_1^* and H_0^*/H_1^* respectively in the 'direction' of the specific ϕ function used. Next we have obtained their asymptotic powers for local alternatives to H_0^*, and have established that if all marginals of F are identical and the location parameters $N^{-1/2}\delta_i^{(\alpha)}$ are the same for all α, then T_1 is asymptotically $\chi^2((p-1)(k-1))$. Note from Theorem 1 that H_1^* is satisfied in such a case.

Computer programs for T_0 and T_1 have been written for specific functions ϕ_V, ϕ_B, ϕ_L and the multivariate version (see [7]) Kruskal-Wallis H-statistic. (It has been noted (see, e.g., [7]) that W-statistics (i.e., T's using ϕ_W) have the same limiting properties as H.) Also simulation studies have been carried out to investigate χ^2 approximations under H_0^* and powers under some alternatives to H_0^* (some satisfying H_1^*) for three different distributions and several covariance structures. These studies are being presented in another paper [4] and these seem to indicate that, apart from the partial justification provided for the test T_1 for the hypothesis H_1^*, there is also reasonable empirical justification to believe that indeed the concept of local alternatives to H_0^* in the direction of H_1^* might indeed provide the way out of the theoretical hurdle encountered earlier.

References

[1] Bhapkar, V.P. (1961). A nonparametric test for the problem of several samples. Ann. Math. Statist. 32, 1108-17.

[2] .... (1965). Some nonparametric tests for the multivariate several sample location problem. Proc. Internat. Symp. Mult. Analysis. Academic Press, New York, 29-42.

[3] .... and Patterson, K.W. On some nonparametric tests for profile analysis of several multivariate samples. University of Kentucky, Department of Statistics Technical Report Series No. 100.

[4] Bhapkar, V.P. and Patterson, K.W. (1976). A Monte Carlo study of some multivariate nonparametric statistics for several samples. University of Kentucky, Department of Statistics Technical Report Series No. 101.

[5] Morrison, D. (1967). Multivariate Statistical Methods. McGraw-Hill, New York.

[6] Puri, M.L. and Sen, P.K. (1971). Nonparametric Methods in Multivariate Analysis. John Wiley and Sons, New York.

[7] Suguira, N. (1965). Multisample and multivariate nonparametric tests based on U-statistics and their asymptotic efficiencies. Osaka Journ. Math. 2, 385-426.

Acknowledgment: Research sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-75-2841. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes not withstanding any copyright notation hereon.

PROBLEMS IN MEASURING EARNINGS FOR BENEFIT COST ANALYSES OF HUMAN INVESTMENT PROGRAMS

John Bishop, Institute for Research on Poverty
University of Wisconsin--Madison

I. Introduction

The standard method of calculating benefit cost ratios or social rates of return of human investment programs involves comparing earnings gains received by program participants to the foregone time cost of their participation and the direct costs of the program. Neoclassical production theory implies a correspondence between such earnings measures and the true contribution to real output of a program if all markets are competitive and taxes, fringes and marginal externalities are all zero. Since these assumptions seldom hold, standard benefit cost calculations will often be biased. This paper explores the practical implications of relaxing the assumption of zero taxes and fringes and of competitive product markets. The more difficult problems posed by externalities and noncompetitive labor markets can only be dealt with in the context of the specific type of human investment involved, therefore these issues are left for another paper.

The true contribution to real output of a marginal increment in a given factor of production (MVP) is the price consumers pay for the product times its marginal physical product. Thus, what we desire to measure is MVP (MVP = P · MPP). The paycheck of the worker and, therefore, his reported earnings (E) can be substantially smaller than this.

A profit-maximizing firm arranges its use of factors so that the total cost of employing a marginal increment of a factor equals that factor's marginal revenue product (MRP = MR · MPP). The paycheck consequently will equal the factor's MRP minus fringe benefit payments and taxes on the factor's use.

$$E = \frac{MRP}{(1+t_s+f)} = \frac{MVP}{(1+t_s+f)} \cdot \frac{MR}{P} = \frac{MVP}{(1+t_s+f)(1+t_e+\pi_m)}$$

where t_s = average rate of employer-paid Social Security tax

f = average rate of fringe benefit contribution

t_e = excise tax rate

π_m = average rate of monopoly profit.

The marginal revenue (MR) is lower than the product's price to the extent that there are excise or value-added taxes on output or less than infinitely elastic firm demand curves.

An additional potential source of error in calculating the social benefits of programs is inaccuracy in the reporting of earnings. In section II each of these five potential sources of discrepancy between MVP and reported earnings (E_r) will be examined and their magnitude estimated. In section III average and marginal rates of tax and fringe benefit payments are calculated for each of six earnings classes. When all five discrepancies are put together, the resulting average MVP/E_r for the 1970 Census declines from 1.15 at low earnings levels to 1.11 at high earnings levels. Section IV examines how these estimates of coverage bias affect the benefit cost ratios and rates of return of human investment programs.

II. Discrepancies between MVP and Reported Earnings (E_r)

Fringe Benefits

Fringe benefits are a large and growing part of labor compensation. As a percentage of total earnings they have risen from 1.1 percent in 1940 to 4.7 percent in 1959 and 6.2 percent in 1969. Some fringes are given disproportionately to high-income individuals--e.g., stock options and pensions. Others such as health insurance, food and housing received as pay, and farm products consumed at home are a larger portion of the compensation of employees with low earnings.

Recent studies by the Social Security Administration of pension and group medical insurance eligibility of a national probability sample of workers and of the pension receipts of recent retirees allow us to make some rough calculations of fringe benefit coverage (Kolodrubetz, 1974; Kolodrubetz and Landay, 1973). Columns 1 and 6 of Table 1 present the proportion of full-time wage and salary workers that are covered by group medical insurance and by private retirement plans in May, 1972. Since private pensions are generally supplemental to Social Security, the need and demand for them is greater among better paid workers. Coverage by retirement plans rises from 26 percent for those earning less than \$5000 to 70 percent for those earning more than \$20,000. Coverage by group health plans also rises with income but is high throughout.

Part-time and self-employed workers are not likely to be covered by group medical or retirement plans. In order to estimate the pattern of coverage for all workers, the estimates of coverage in columns 1 and 6 have been revised downward to take into account the incidence of part-time and self-employed workers by income group.

The next step is to estimate for eligible workers the relationship between the individual's earnings and the amount of fringe benefit costs incurred on his behalf by his employer. The Survey of New Beneficiaries (Kolodrubetz, 1973) recently completed by the Social Security Administration has found that the median ratio of pension to earnings on an employee's longest job follows a gentle U-shaped pattern as he moves up the earnings distribution (column 8). The increase in this ratio that occurs as the individual's earnings rise from \$5000 to \$15,000 is quite gentle, however, and is due to the greater length of service of the higher earnings retirees.

Unlike pensions, whose costs are roughly proportional to earnings, medical insurance costs should be more or less constant across individuals. On the one hand, government and private establishments with high wage structures tend to have the most comprehensive insurance coverage. On the other hand, people in low wage occupations are more likely to use the coverage they have so the cost of their insurance may be higher. The net effect of these counteracting influences is assumed to produce an elasticity of group medical insurance expenditures to earnings of about one-third. Dollar amounts of group health insurance per covered employee were, therefore, assumed to rise one dollar for every 100 dollars of income.

The aggregate amount of employer contributions to employee benefit plans is obtained by summing National Income Account estimates of other labor income, federal employee pensions, state employee pensions and one-half of defense health expenditures. Pensions are approximately one-half the total, and medical insurance approximately one-third. The rest is made up of life insurance and temporary disability plans. This category is distributed among workers in proportion to earnings. Estimates of the total amounts of and distribution of food and housing expenditures received as wages are taken from a study by Herriot and Miller (1972). Line 2 of Table 2 presents rates per dollar of earnings. After calculating average fringe benefit estimates for workers at their specific earnings level, the marginal rates of fringe compensation were calculated by relating the difference between these averages to the earnings increment.

Taxes on Labor Input (t_s)

The second source of discrepancy between reported earnings and the marginal revenue product are the Social Security and Unemployment Insurance taxes paid by employers. The effective average rate of tax for earnings brackets below the maximum taxable wage is equal to the proportion of workers covered (.873) times the statutory rate. In 1969 the Social Security tax was paid on the first $7800 of wages, so for incomes above $7800 average tax rates are the maximum tax, $327, times .873 divided by the midpoint of the earnings interval. The unemployment insurance tax rate (.0138) and maximum taxable wage ($3400) is a weighted average of varying state provisions.

Excise Taxes

State and federal sales and excise taxes, not including alcohol and tobacco taxes, totaled $27.6 billion in 1968. Alcohol and tobacco taxes are excluded because they are assumed to reflect the negative externalities one's use of these products imposes on others.

A separate calculation was carried out for retail gasoline taxes (9.187 billion in 1970) in order to account for their greater incidence on imports and property income. With backward shifting the 1967 input-output table implies that the burden falls 3.4 percent on imports, 26.8 percent on the property income of petroleum extraction, refining, wholesaling and retailing, 19.4 percent on employee compensation of these industries and 50.4 percent on purchases from other industries. In 1969, the ratio of fuel taxes assigned to oil industry employee compensation to economy-wide employee compensation is .00315. The ratio of general and specific sales taxes (including .504 of the gas tax) to GNP was .0294. The incidence of excise taxes on labor compensation is, therefore, .03255.

Monopoly Power

Whether monopoly power makes a further correction desirable depends upon the source of the monopoly and which factor of production is receiving the monopoly rents. If monopoly rents add equal percentage increments to workers' wages and to capital's return, no problem arises, for our compensation data has already captured them. If a firm faces an infinitely elastic long-run demand curve at its limit price, but nevertheless receives monopoly rents because of the ownership of some unique factor of production (e.g., patents, control of the best raw material sources, government licenses), no adjustment is required. An add-on is required only where $P > LRMR = LRMC$ and where the monopoly rents do not get paid to labor.

How large might such monopoly profits be? Harberger's (1954) upper bound estimate of the welfare impact of monopoly implies that one-third of manufacturing profits are excess profits. Assuming that the share of monopoly rents $[(P - LRMC)q]$ in corporate profits is one-third for manufacturing, we obtain an upper bound estimate of $17.7 billion for 1969, or 1.9 percent of GNP. The results presented in Tables 2 and 3 do not include an adjustment for monopoly distortions or for systematic economies of scale. The reader may make his own adjustment for monopoly with his own assumption about monopoly by simply multiplying the average and marginal ratios of social benefit to reported income in Table 3 by a number between 1 and 1.019.

Underreporting

It is possible to determine the average degree of under- or overreporting by Census interviewees for each type of income by comparing national income aggregates derived from establishment sources with the aggregates implied by the Census household data. While 96 percent of wage and salary income was reported in the Current Population Survey, only 52 percent of farm income was reported (Projector and Brety, 1972). The percent of aggregate earnings missed was only 1.5 percent in the 1970 Census, .67 of a percent in the 1960 Census, and 5.4 percent in the 1970 and 1971 CPSs.

Census and CPS aggregates may be low either because people are missed or because on average each person understates his earnings. Only the latter source of discrepancy will cause a bias in measures of benefits of human investment programs. The Bureau of the Census has developed estimates of the amount by which the nation's population was understated in the Census and CPS (Siegel, 1967 and 1974). By applying age, sex, and race-specific undercount rates to 1970 Census estimates of their earnings aggregates, and assuming that those not enumerated earn two-thirds the average, we can estimate the effect of the Census undercount on earnings aggregates. The undercount adjustments of Census aggregates were 2.14 percent in 1960 and 2.8 percent in 1970. This implies in turn that per capita earnings in the Census overstated the true level of per capita earnings by 1.47 (2.14-.67) percent in 1960 and 1.34 percent in 1970. The CPS understated true per capita earnings by 2.8 percent.

III. Combining the Estimates

In Table 2 we collect our estimates of earnings-bracket-specific correction factors. We find that the average social productivity benefit of a person's work--the sum of after-tax earnings and taxes generated--averages about 113 percent of reported earnings. As earnings rise, the ratio of social benefit to census-reported earnings tends to fall from 1.15 to 1.11. The fall in the ratio is a consequence of imputations not rising as fast as income and the zero marginal Social Security tax on wages above $7800.

Dividing the Social Return into Private and Public Components

What portion of this total or social return can the individual be expected to take into account when he makes his own decisions? Splitting the social return into private and public components is necessarily more arbitrary than calculating the total return. Lines 9 and 14 of Table 2 present lower bound estimates of the ratio of private benefits to reported earnings. It is based upon the assumption that extra earnings do not, on the margin, place any additional burden on the government's provision of services. This is a valid assumption for pure public goods such as defense, foreign affairs, space, and police and fire protection. Providing an individual with more of a pure public good inevitably means everyone else gets more.

However, for many government services provided at zero or nominal cost, giving the service to one person means it must be denied to someone else. If usage of such services rises with income, extra after-tax income places an additional burden on other taxpayers. Some directly provided services of this kind are education, libraries, airports, congested high ways, recreation, sewers, water supply, and garbage collection. From a life cycle perspective the largest of the transfer programs, Social Security, also provides larger dollar benefits to people with higher earnings. Usage of certain other services--Food Stamps, directly subsidized housing, Medicaid, unemployment insurance and AFDC--go down as earnings rise. Studies that lump all these effects together obtain small (about 4 percent) positive marginal effects of earnings on net usage of government services (Reynolds and Smolensky, 1974). Consequently, assuming marginal induced government services to be zero places a lower bound on the private benefit. Our estimate of the private benefit to reported earnings ratio is thus simply the MVP/E_r ratio minus the average rate of personal, excise and labor input taxation.

In 1970 the average rate of tax for the income tax and Social Security (employee share) together rises from 7.9 percent of earnings in the $2000-$4000 bracket to 17.2 percent in the $15,000-$25,000 bracket. The ratio of after-tax compensation to reported earnings, therefore, falls from .98 to .89 as one moves from low to high brackets. The ratio of all taxes generated to reported earnings rises from .17 to .22 as earnings rise.

Marginal rates of tax are higher, ranging from .26 to .21. Marginal ratios of MVP and after-tax compensation to earnings are required because the effect of most human investment programs is a rise in the individual's earnings. After-tax income and total tax generated are calculated for the representative family in each bracket using average ratios and the midpoints of the intervals as family income. The difference between the predicted figure for adjacent income brackets is divided by the rise in income from the midpoint of one bracket to the next to obtain marginal ratios. The marginal ratio of after-tax compensation to reported earnings is below the average ratios and tends to fall with income from a high of .935 to .87. The marginal ratio of social or total return to reported earnings is also below the average, falling from 1.12 in low brackets to 1.08 in high brackets.

IV. The Impact of Coverage Bias on Benefit-Cost Ratios and Rates of Return

Benefit-cost ratios and rates of return express a relationship between benefits received in the future and costs incurred now. The net effect of coverage bias on a benefit-cost ratio or rate of return depends upon the relative size of the bias in measuring each component.

In most benefit-cost or rate of return calculations, benefits are assumed to accrue to individuals as gains in earnings. We have seen that traditional measures of these social benefits understate benefits by 8 to 15 percent. The coverage bias in measures of cost varies even more from situation to situation for there are three distinct types of cost: government budgetary costs, forgone work time costs and forgone leisure time. Each will be discussed in turn.

Net Coverage Bias When all Costs are Budgetary

The coverage bias issue has a counterpart in the measurement of budgetary cost. Correct measurement of a program's cost requires the inclusion of pension cost and other fringe benefits being earned by individuals assigned to the program. This occurs as a matter of course when the task or service is provided by an independent agency on a contract basis. Separate calculation of an appropriate fringe benefit rate may be necessary when government-wide pension programs are financed on a pay as you go basis. Accounting procedures for allocating fringe benefit costs to programs are already well developed so we will assume that budgetary costs are correctly measured (i.e., fringes are included).

The net coverage bias in the benefit-cost ratio of programs which do not require a time input on the part of the beneficiary is equal to the coverage bias inherent in the particular benefit being analyzed. The GNP benefit of a health program that reduces mortality is the marginal value product of the workers whose deaths are averted. If these workers earn $20,000, the true GNP effect is 10.8 percent greater than the earnings loss. (The coverage bias ratio = MVP/E_r = 1.108.) If the program averts the death of workers earnings $3000, traditional benefit measures understate the GNP effect by 14.8 percent. The degree of understatement is greater for health investments in low-wage workers.

In most cases a human investment program produces marginal increases in the earnings of individuals. Under these circumstances, the appropriate coverage bias ratio is $\Delta MVP/\Delta E_r$. It can be found on line thirteen of Table 2 and in the first three columns of Table 3. An improvement in the quality of education, or training, or a health program that deals with a nonfatal disease are examples of such a program. Traditional measures of the benefits of a training program targeted at low-wage workers have a coverage bias ratio of between 1.12 and 1.129. The coverage bias ratio for improved graduate education is 1.09. As with mortality reduction programs, the understatement is greatest for low-wage workers.

Net Coverage Bias When Forgone Work Time Is Part of the Cost

In most training programs a major part of the social cost is the work time sacrificed by the trainee. Decisions to expand the number of people attending college or receiving training necessarily imply reductions in labor supply during the training period. Measures of earnings forgone because of the reduction in labor supply are subject to the same type of coverage bias as benefit measures. Since the loss of earnings during the training period generally occurs when the individual is young and has a low earnings capacity, the coverage bias ratio of this cost element is typically larger than the ratio for the corresponding benefit. The adjustment factor for forgone earnings costs of schooling are given in columns 7-9 of Table 3. The social costs of schooling include, however, a government expense component that does not suffer coverage bias. The coverage bias of total costs depends upon the relative importance of these two elements. Instructional costs of the first two years of college are approximately .37 of total costs, so the coverage bias ratio for the costs of the first two years of college is roughly 1.076 (.63 (1.12) + .37 (1.0) using 1970 Census data. The net coverage bias ratio is thus 1.023 (1.10/1.076). For this and other schooling increments in which forgone wages are more than half the social cost, the coverage bias in our measure of cost almost exactly counterbalances the coverage bias in the benefit measures.

Net Coverage Bias When Time Costs Are Leisure Forgone

In many instances the time costs of schooling come wholly or partly at the expense of leisure. The 3.4 million part-time students in degree credit college classes are generally working full-time as well. A substantial portion of a full-time student's studying and class time involves a reduction of leisure. Using National Longitudinal Survey data, Parsons (1974) has estimated that the share of leisure time in the 1300 hours required for full-time school attendance is 52 percent for 17 year olds, 34 percent for 19 year olds, and 21 percent for 21 year olds.

The social cost of a sacrificed hour of leisure time is not as great as the social cost of a sacrificed hour of work. The difference is in the tax revenue produced by the work. Young people adjust their hours of work until, on the margin, they receive approximately equal satisfaction from extra leisure and from extra work. The dollar value of leisure time is, therefore, roughly equal to the after-tax wage rate. If we have valued forgone leisure time at the money wage rate, we have exaggerated its cost. The ratio of the social value of leisure to the reported wage rate is equal to ratio of the private value of work to reported earnings. These ratios, the combined tax and coverage bias when a benefit or cost is nonpecuniary, are given in columns 4-6 and 10-12 of Table 3.

If a human investment program is structured so that time inputs are forgone leisure, net coverage bias is quite large. In the junior college example dealt with above, tax and coverage bias in our estimate of cost is .63 (.9) + .37 (1.0), or .937. Net coverage bias in the benefit cost ratio is 1.174 (1.10/.937) when all time inputs result in a sacrifice of leisure.

The net coverage bias is also easy to calculate when both work and leisure have been reduced. For a 19 year old full-time student the coverage bias in the time cost is .34 (.9) + .66 (1.12) or 1.045. Net coverage bias in the cost-benefit ratio for full-time attendance of a 19 year old in junior college is, therefore, 1.10/(.63 (1.045) + .37 (1.0)) or 1.070. Taking into account the fact that part of the time input of schooling results in a sacrifice of leisure and not of work raises the magnitude of the net coverage bias, especially of young students. Adopting Parson's estimates of work's share of time inputs, our new estimate of net coverage bias in benefit-cost ratios for additional students is 1.131 for 9th and 10th grade, 1.103 for the last two years of high school, 1.057 for the last two years of college, and 1.031 for graduate school. These differentials in the coverage bias reflect the fact that the social efficiency of a human investment is sensitive to whether the time invested comes at the expense of work or of leisure.

Net Coverage Bias in Rates of Returns

The net coverage bias in rates of return can be calculated by a slight modification of the procedure used in our examples. Adjustment factors are applied separately to costs and benefits, the present value of costs are set equal to the present value of benefits, and the equation is solved for the internal rate of return. The change in the rate of return that the coverage bias adjustment produces is similar to the net coverage bias of the corresponding benefit-cost ratio. If there is no time variation in the level of benefits or costs, the ratio of the new to old rate of return will equal the net coverage bias ratio, the correction factor for benefit-cost ratios. Since, however, the dollar size of the benefit tends to rise with age, the proportionate change in the rate of return produced by the coverage bias adjustment will tend to be smaller than the proportionate change of benefit-cost ratios.

References

Harberger, Arnold. 1954. "Monopoly and resource allocation." American Economic Review 44 (May 1954):77-86.

Herriot, R. and Miller, Herman. 1972. "Who paid the taxes in 1968." Mimeographed. Census Bureau.

Kolodrubetz, Walter. 1973. "Private retirement benefits and their relationship to earnings; survey of new beneficiaries." Social Security Bulletin 36 (May 1973):16-36.

Kolodrubetz, Walter. 1974. "Group health insurance of full time employees, 1972." Social Security Bulletin 37 (April 1974): 17-34.

Kolodrubetz, W. and Landay D. 1973. "Coverage and vesting of full-time employees under private retirement plans." Social Security Bulletin 36 (Nov. 1973):20-36.

Parsons, D.O. 1974. "The cost of school time, foregone earnings, and human capital formation." Journal of Political Economy 82 (March/April 1974):251-266.

Projector, D. and Brety, J. 1972. "Measurement of transfer income in the Current Population Survey." Conference on Research in Income and Wealth, October, 1972. National Bureau of Economic Research.

Reynolds, M. and Smolensky, E. 1974. "The post fisc distribution: 1961 and 1970 compared." National Tax Journal 27 (Dec. 1974):515-30.

Siegel, J. 1967. "Completeness of coverage of the nonwhite population in the 1960 Census and current estimates, and some implications," in Social Statistics and the City, ed. David Heer. Report of June 1967 Conference, Joint Center for Urban Studies of MIT and Harvard.

Siegel, J. 1974. "Estimates of coverage of the population by sex, race and age in the 1970 Census." Demography 11 (Feb. 1974).

Table 1

Calculation of Marginal Fringe Benefit Rates

Earnings	Group Health Insurance					Retirement					Total	
Men	% FTWS with group health[1]	% all with group health (est)[2]	Avg cost per emp (est)[3]	Avg % of earnings	Marg % of earnings	% FTWS covered by ret. plan[4]	% all workers covered[5] (est)	Median pension earnings ratio[6]	Avg % of earnings	Marg % of earnings	Marg fringe benefit	Increment in earnings (000's)
less than 5000	.61	.53	202	3.57		.26	.17*	.30	.98		5.66	0→3
					1.8							
5-6000	.75	.69	222	2.78				.23			6.10	3→5
					2.5					3.19		
6-7000	.82	.77	232	2.75				.22			6.50	5→7
					2.0	.58	.50*		2.31			
7-8000	.86	.82	242	2.65				.24			7.90	7→9
					2.1							
8-9000	.91	.87	252	2.58				.25			7.75	9→12.5
					.9					5.64		
9-10000	.91	.87	262	2.40				.25				
					1.1							
10-15000	.93	.89	292	2.08		.74	.70*	.27	3.64			
					.9					4.50	6.49	12.5→17.5
15-20000	.93	.89	342	1.74		.76	.72*	.28	3.89			
					.9					3.89	5.88	17.5→30
more than 20000	.92	.89				.70	.72*	.28	3.89			
All FT	.80					.47		.25				
Avg All			252	2.11					3.22		6.49	
Women												
											5.27	0→3
less than 5000	.59	.50	202	3.36		.31	.22	.19	.80			
					1.76					2.16	5.03	3→5.5
5-6000	.77	.70	222	2.82		.58	.44	.17	1.44			
					1.61					4.02	6.74	5.5→8
6-10,000	.81	.76	257	2.44			.51	.23	2.26			
					.95					2.54	4.60	8→15
more than 10,000	.84	.80	327	1.74		.60	.54	.23	2.39			

Table 2

Incidence of Overreporting, Fringe Benefits, and Taxes by Income Class in 1969

Family Income (in thousands)	2-4	4-6	6-8	8-10	10-15	15-25
Average Percent of Earnings						
Under or overreported	-1.34	-1.34	-1.34	-1.34	-1.34	-1.34
Food and housing received as wages	1.69	.82	.57	.43	.32	.21
Health and Pensions	5.66	5.84	6.03	6.44	6.82	6.62
Social Security tax on employer	4.19	4.19	4.19	3.64	2.62	1.64
Unemployment insurance tax	1.0	.70	.50	.38	.28	.17
Personal income tax	3.67	6.45	8.64	10.12	12.21	14.54
Excise Tax % of Compensation	3.26	3.26	3.26	3.26	3.26	3.26
Ratio of MVP to E_r	114.8	113.8	113.5	113.1	112.2	110.8
Ratio of after-tax compensation to E_r	97.6	94.1	91.9	91.3	90.7	89.1

Earning Increments (in thousands of dollars)	1→3	3→5	5→7	7→9	9→12.5	12.5→20
Impacts of Earning Increments (percent)						
Marginal fringe benefits	4.66	5.6	6.5	7.90	7.75	6.5
Employer-paid taxes on labor input at margin	5.19	4.39	4.19	1.68	0	0
Marginal income tax	5.00	11.12	14.13	15.30	17.60	18.40
ΔMVP / ΔRepted earnings	112.0	112.2	112.9	111.8	109.9	108.6
ΔAfter-tax compensation / ΔRepted earnings	93.6	88.4	86.3	89.4	88.8	86.8

Table 3

Coverage Bias in Traditional Measures of the Benefits and Costs of Education (Ratios of True Productivity Benefits or Costs to Reported Money Earnings for Males)[1]

	Productivity Benefits						Student's Time Costs					
	Social			Private			Social[2]			Private[3]		
	Census 1959	Census 1969	CPS	Census 1959	Census 1969	CPS	Census 1959	Census 1969	CPS	Census 1959	Census 1969	CPS
Finish Elementary[4]	1.08	1.12	1.16	.87	.88	.92	1.11	1.15	1.20	1.02	1.01	1.05
Elem. to HSDO[4]	1.08	1.12	1.16	.87	.88	.92	1.11	1.15	1.20	.95	.94	.98
HSG	1.08	1.11	1.15	.87	.89	.93	1.08	1.13	1.17	.93	.92	.96
Coll. Dropout	1.07	1.10	1.14	.87	.89	.93	1.08	1.12	1.16	.91	.90	.94
Coll. Grad.	1.07	1.10	1.14	.87	.88	.92	1.08	1.12	1.16	.90	.89	.93
Grad. School	1.06	1.09	1.13	.85	.87	.91	1.08	1.12	1.16	.89	.88	.92

[1] By a simple manipulation of these factors, rates of return and benefit-cost ratios may be adjusted for taxation and coverage bias. Calculate the cost adjustment by taking a weighted average of the student time cost factor and one. The weights are the conventionally calculated foregone earnings and either instructional cost or out of pocket tuition and book costs. This average is divided into the productivity benefit adjustment factor. Female productivity benefit ratios will tend to be lower because of lower fringe benefits and higher income taxes and higher because of the greater relative importance of Social Security taxes.

[2] If time spent in schooling would have been spent working.

[3] The social cost if time at school comes at the expense of leisure or the private cost no matter how the school time would have been spent.

[4] Foregone time costs adjustment for students through 10th grade use the average ratio rather than the marginal ratio that is assumed for all other levels of schooling. In other words, until the 10th grade, it is assumed that those in school hardly earn anything at all. For all others, it is assumed those in school already work some and that the effect of dropping out is to increase the amount of work and leisure from an already existing base and that therefore, marginal rates of taxation and coverage bias apply. It is further assumed that elementary students pay only Social Security taxes on their earnings.

CLASSIFICATION BASED ON INCOMPLETE DATA RECORDS

Tom Bohannon and W. B. Smith, Texas A.& M. University

Introduction

In this paper we discuss the situation where there are two distinct multivariate normal populations, Π_1 and Π_2, with common variance-covariance matrix. We observe a p-vector X and must assign X to Π_1 or Π_2 based on the components of X and our classification rule. If the parameters of the distributions of populations Π_1 and Π_2 are known, this information is utilized in the construction of a classification rule. If the parameters are not known, which is the usual situation, then random samples from Π_1 and Π_2 are used to estimate these parameters and to construct a classification rule. We shall use as our classification rule, Anderson's discriminant function.

One of the problems that arise in the practical applications of discriminant analysis is analytically measuring the goodness of the classification rule. This rule must be evaluated based on some criterion of goodness of classification. For our criterion in this study we shall use the total probability of misclassification. In general, since the parameters of the populations are usually unknown, the probability of misclassification must also be estimated from random samples.

The samples on which one bases the classification rule and estimates the probability of misclassification often contain incomplete observation vectors, that is, vectors in which one or more components are missing. In many such situations these incomplete vectors are not included in the construction of a classification rule or in the estimation of the probability of misclassification. The primary purpose of this paper is to investigate a method for incorporating these incomplete observation vectors in the construction of the classification rule and the estimation of the probability of miscalssification. This method and the commonly practiced method of ignoring these incomplete vectors will be compared by computer simulation.

The use of discriminant analysis techniques on incomplete data sets is an area where very little research has been done. Jackson (1968) investigated a classification problem which had missing values in a large data set. The missing values were estimated using means and regression techniques and for the problem under study, the estimation procedure using missing data gave better results than the procedure of ignoring the observations with missing values.

Chan and Dunn (1972) investigated the problem of constructing a discriminant function based on samples, which contained incomplete observation vectors. Several methods of estimating the missing components of these vectors were utilized and the resulting vectors were used to construct the discriminant function. They concluded that no method was best for every situation, and gave guidelines to use in choosing the best method for various situations.

Hocking-Smith Estimation Procedure

A generalization of the estimation procedure reported by Hocking-Smith (1968) will be applied to random samples from multivariate normal populations, which contain incomplete vectors. This procedure requires that optimal estimators of the mean vectors and dispersion matrices are available for each group of observations, the groups being collections of observation vectors with identical patterns of incompleteness. The procedure has been shown to be essentially equivalent to solving the maximum likelihood equations for the incomplete situation. The estimators have been shown to be consistent and asymptotically efficient.

Note that we are estimating the mean vectors and the variance-covariance matrices without estimating the missing components of the incomplete vectors. Those missing components could, however, be estimated by using the previously mentioned estimators and regression techniques.

To illustrate the form of the estimators, consider a set of observations which follow a p-variate normal distribution with unknown mean vector μ_1 and variance-covariance matrix Σ_1. Let there be n_1 independent complete observation vectors and n_2 independent incomplete observation vectors which follow the q-variate marginal distribution. We define the elementary matrix D such that $\mu_2 = D\mu_1$ is the mean of the marginal and $D\Sigma_1 D' = \Sigma_2$ the variance-covariance matrix for the marginal distribution.

The joint likelihood L for these two groups of observations is given by $L = L_1 \cdot L_2$, where L_1 and L_2 are the likelihood functions associated with the two groups of observation vectors. The Hocking-Smith estimates are given by

$$ {}_2\hat{\mu}_1 = \hat{\mu}_1 - \frac{n_2}{N} \hat{R}'(D\hat{\mu}_1 - \hat{\mu}_2) $$

$$ {}_2\hat{\Sigma}_1 = \hat{\Sigma}_1 - \frac{n_2}{N} \hat{R}'(D\hat{\Sigma}_1 D' - \hat{\Sigma}_2)\hat{R} $$

where $\hat{R}' = \hat{\Sigma}_1 D'(D\hat{\Sigma}_1 D')^{-1}$ and $N = n_1 + n_2$. These estimates are in general maximum likelihood considering the combined likelihood function L if $\hat{\Sigma}_2$ is replaced by $\hat{\Sigma}_2 + \hat{H}_2$,

where $\hat{H}_2 = n_2(\hat{\mu}_2 - D_2\hat{\mu}_1)(\hat{\mu}_2 - D_2\hat{\mu}_1)'$.

The extension to more than two groups follows sequentially and is easily adaptable to computer programming. For further information the interested reader is referred to Hocking *et al.* (1969).

Estimators of the Probability of Misclassification

The problem of estimating the probability of misclassification has received a considerable amount of attention in the statistical literature. A fairly complete review of the literature on this problem is given by Toussaint (1974). The estimators that are considered in this study are as follows:

1. The estimator $\phi(-D/2)$, where
   $D^2 = (\hat{\mu}_1 - \hat{\mu}_2)' \hat{\Sigma}^{-1} (\hat{\mu}_1 - \hat{\mu}_2)$.
2. The estimator $\phi(-D'/2)$, where

   $D'^2 = (n_1 + n_2 - p - 3)(n_1 + n_2 - 2)^{-1} D^2$.

3. The McLachlan estimator, which is defined in

McLachlan (1975).

For the discriminant functions based on only the complete observations these estimators will be denoted by P_1, P_2 and P_m, respectively. For the discrominant functions based on all of the observation vectors these estimates will be denoted by $\hat{\alpha}_1, \hat{\alpha}_2$ and $\hat{\alpha}_m$, respectively. The estimators $\hat{p}_1$ and $\hat{p}_2$ were studied by Lachenbruch and Mickey (1968) and Sorum (1972).

<u>Simulation Procedure and Results</u>

Random samples are generated from each of two populations and a specified percentage of vectors are randomly chosen and made incomplete. The groups for the Hocking-Smith estimator procedure are formed and the estimates of the parameters for each group calculated. The estimates from the group of complete vectors are utilized in the calculation of the discriminant function and the previously mentioned estimators for the probability of misclassification are calculated. Next, the Hocking-Smith estimates are used in calculating the discriminant function and the estimators for probability of misclassification are calculated. This procedure is repeated at least ten times for each specified set of simulation variables. The mean and standard deviation are calculated for each estimator of the probability of misclassification.

The simulation results obtained by varying the values of the correlation coefficient ρ from 0 to .9 by increments of .1 indicated that the relative performance of the estimators were not effected by this simulations variable. The same was found to be true for varying the form of the mean vector. Hence the simulation results presented in the tables and in the figures are for the pooled simulations of ρ and the form of the mean vector. The tables and figures presented in this paper are only for the number of variables equal to three and the Mahalanobis distance equal to four. The percentage of missing values were chosen to be 20, 40 and 80 with three groups of vectors for the Hocking-Smith estimation procedure. These simulation results are part of the simulations conducted by Bohannon (1976) and are representative of those results.

Frequencies and cumulative proportions of

$$e = |\alpha - \hat{\alpha}|$$

where α is the optimum probability of misclassification and $\hat{\alpha}$ as the estimator of this probability were calculated for the simulation combinations with end points .0125, .025, .0375, .05, .0625, .075, and the last interval greater than .075. Figures 1 to 3 present these results and Table 1 gives the means and standard deviations for these estimators.

In analyzing the previously mentioned tables and figures, there are several observations that are apparent. One being, that as the percentage of incomplete data increases, the variances of the estimators increase. However, the increase for the estimators based on the Hocking-Smith estimates is not as great as that for the estimators based only on complete vectors. Our simulations also indicate that the McLachlan estimator has a larger variance in general than the other estimators for our range of population parameters. The simulations indicate that $\hat{\alpha}_2$ is the best estimator of α based on the criterion of unbiasedness and minimum variance and in general the incomplete vectors do provide useful information for classifying the observation vectors.

References

Bohannon, Tom R. [1976]. "Discriminant Analysis With Missing Data." Ph.D. dissertation, Texas A.& M. University, College Station, Texas.

Chan, L. S. and Dunn, O. J. [1972]. "The treatment of missing values in discriminant analysis - I. The sampling experiment." <u>J. Amer. Statist. Ass. 67</u>, 473-477.

Hocking, R. R. and Smith, W. B. [1968]. "Estimation of parameters in the multivariate normal distribution with missing observations." <u>J. Amer. Statist. Ass. 63</u>, 159-173.

Hocking, R. R., Smith, W. B., Waldron, B. R. and Oxspring, H. H. [1969]. "Estimation of parameters with incomplete data." Themis Report No. 12, Institute of Statistics, Texas A&M University, College Station, Texas.

Jackson, E. L. [1968]. "Missing values in linear multiple discriminant analysis." <u>Biometrics 24</u>, 835-844.

Lachenbruch, P. A. and Mickey, M. R. [1968]. "Estimation of error rates in discriminant analysis." <u>Technometrics</u> 10, 1-11.

McLachlan, G. J. [1975]. "Confidence intervals for the conditional probability of misallocation in discriminant analysis." <u>Biometrics 31</u>, 161-167.

Rulon, Phillip J. [1951]. "Distinctions between discriminant and regression analysis and a geometric interpretation of the discriminant function." <u>Harvard Educational Review 21</u>, 80-90.

Sorum, M. [1972]. "Estimating the expected and the optimal probabilities of misclassification." <u>Technometrics 14</u>, 935-943.

Toussaint, G. T. [1974]. "Bibliography on estimation of misclassification." <u>IEEE Transactions on Information Theory IT-20</u>, 472-479.

TABLE 1

TABULATION FOR ESTIMATORS WITH p = 3, Δ = 2, AND α = .1587

			P_1	P_2	P_m	α_1	α_2	α_m
M	20%	mean	.1516	.1589	.1653	.1550	.1625	.1660
ND	126	std. dev.	.0337	.0337	.0357	.0337	.0318	.0332
M	40%	mean	.1462	.1591	.1642	.1513	.1588	.1621
ND	125	std. dev.	.0393	.0393	.0424	.0357	.0357	.0374
M	80%	mean	.1345	.1804	.1908	.1518	.1591	.1626
ND	115	std. dev.	.0580	.0588	.0743	.0409	.0411	.0480

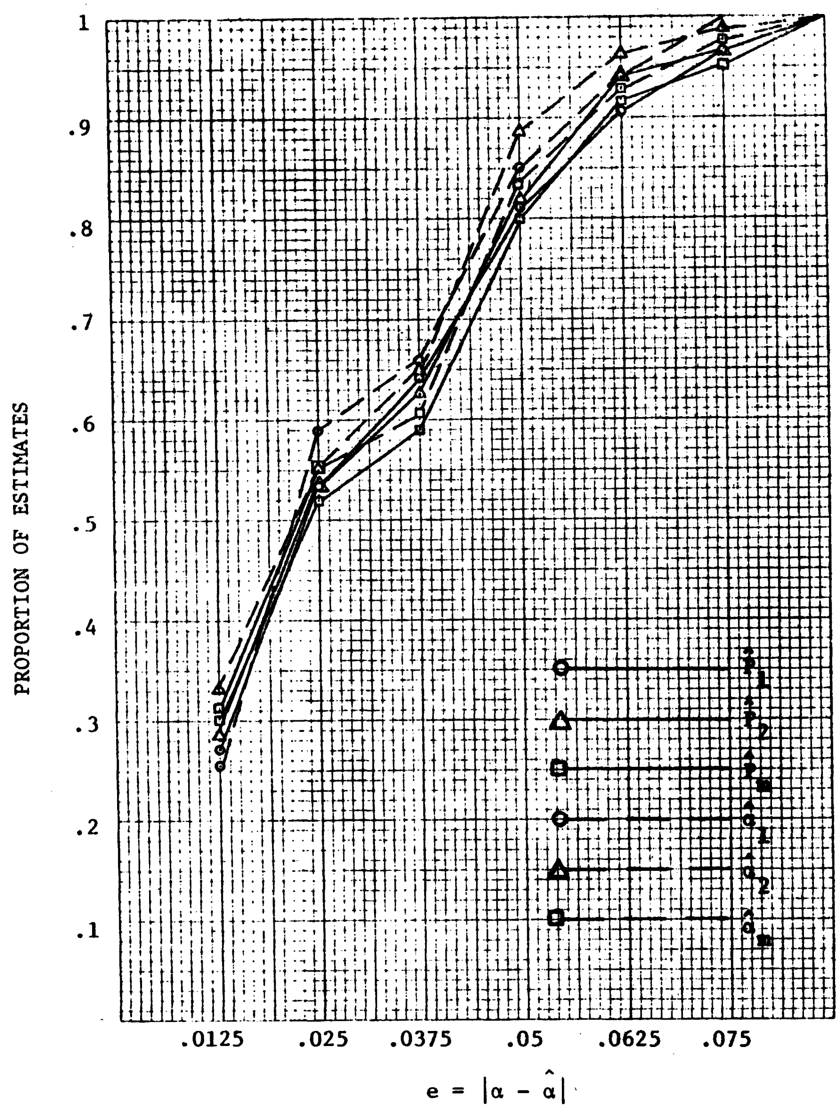


FIGURE 1

126 LINEAR DISCRIMINANT FUNCTIONS WITH $p = 3$, $\Delta = 2$, AND $m = 20\%$

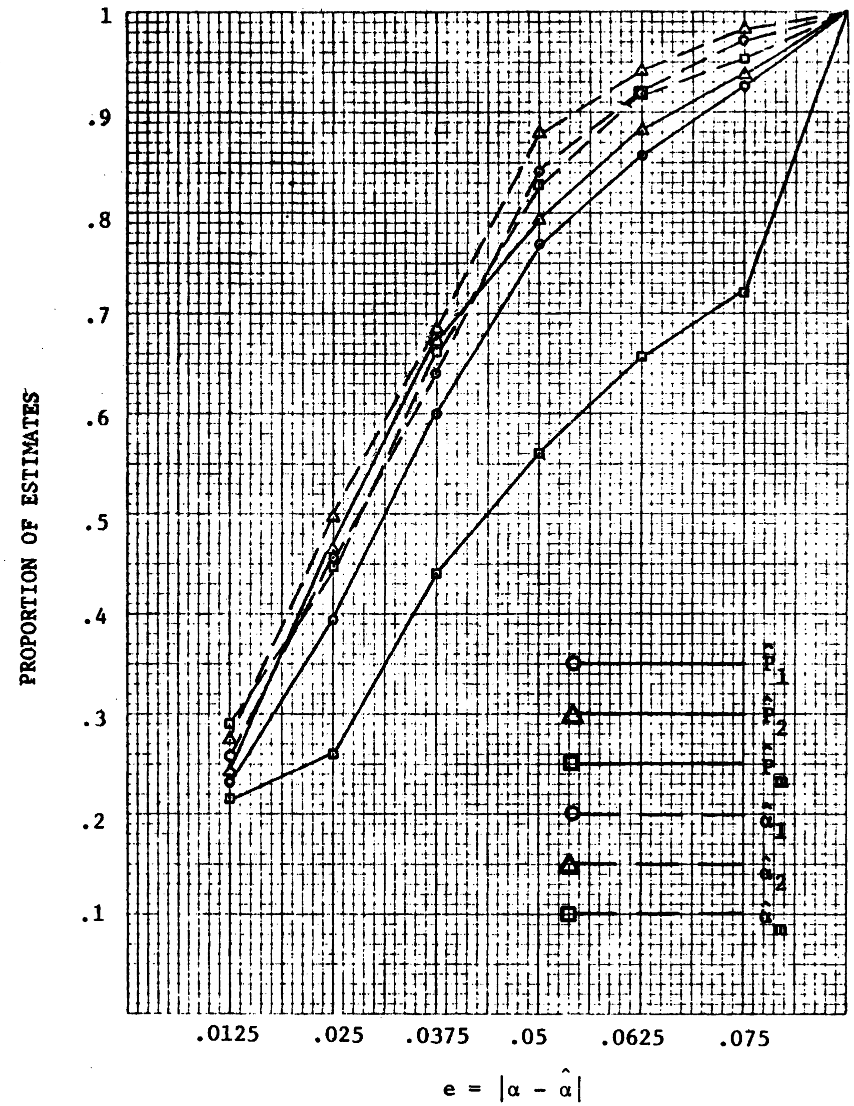


FIGURE 2

125 LINEAR DISCRIMINANT FUNCTIONS WITH $p = 3$, $\Delta = 2$, AND $m = 40\%$

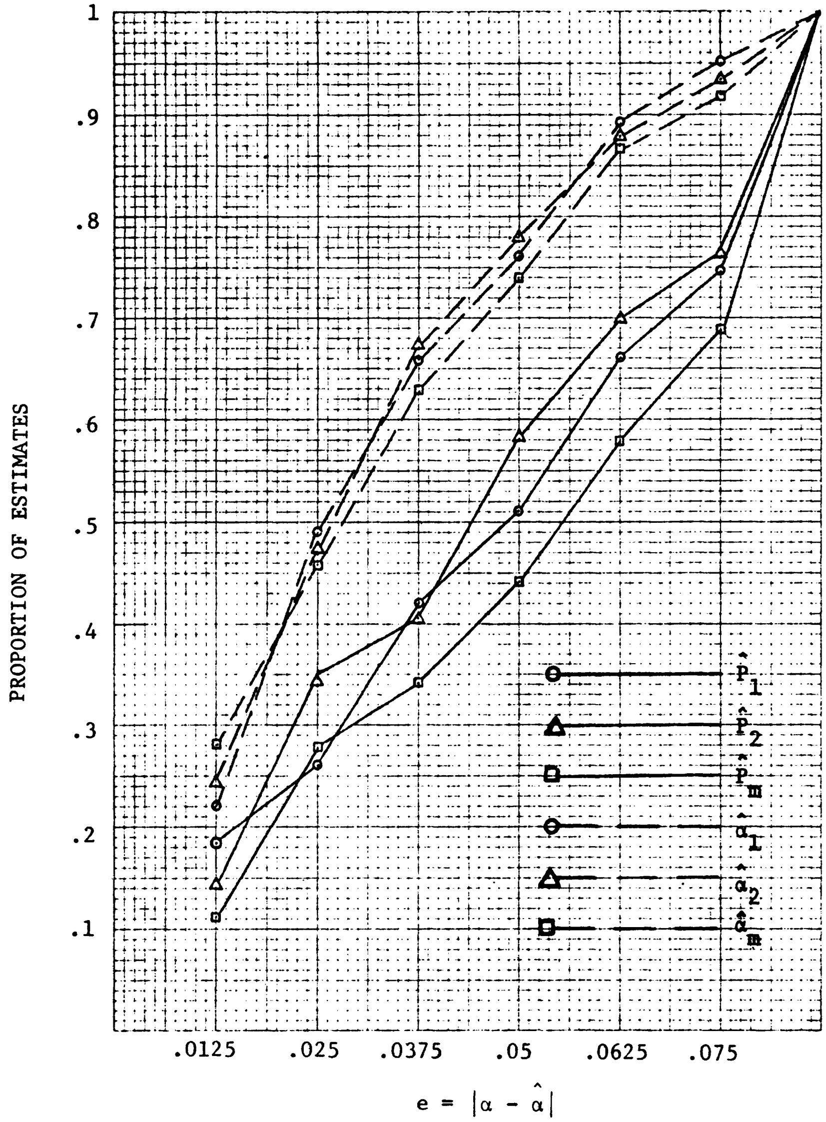


FIGURE 3

115 LINEAR DISCRIMINANT FUNCTIONS WITH $p = 3$, $\Delta = 2$, AND $m = 80\%$

MULTIPLE FRAME ESTIMATION WITH STRATIFIED OVERLAP DOMAIN

Raymond R. Bosecker and Barry L. Ford, US Department of Agriculture

The multiple frame estimator developed by Hartley [1] is extended to take advantage of stratification within the overlap domain for two frame estimation. The overlap domain strata are defined by a stratified list sampling frame. Matching area frame units against list units enables each stratum to be estimated by both frames. A vector of optimum weights combines the stratum estimates from the two frames. The variance of the proposed estimator is shown to be equal to or smaller than the variance of the Hartley estimator. An example using survey data compares the estimate and variance of the proposed estimator to those of other estimators not utilizing stratification.

1. INTRODUCTION

The multiple frame procedure is a common tool of the survey statistician. To estimate a specified variable, random samples are drawn from two or more sampling frames. Most multiple frame surveys involve only two frames:

1. an area frame where the sample unit is a segment of land.

2. a list frame where the sample unit is a name on a list.

Each frame has its own advantages. The list frame is often easier and cheaper to apply. Therefore, *for the same cost*, an estimate using only the list frame generally has a smaller sampling error than an estimate using only the area frame. However, the list frame rarely covers the whole population while the area frame usually does cover the whole population.

The subpopulation which is covered by the area frame but not the list frame is called the "nonoverlap" domain. The remaining part of the population, which is covered by both frames, is called the "overlap" domain. By combining the two estimates of the "overlap" domain, a single multiple frame estimator may be obtained using both the area and list frames. From work done by Hartley [1], the multiple frame estimator of a total, $\hat{Y}$, may be expressed by:

$$(1)\quad \hat{Y}_H = \hat{Y}_{nol} + q\hat{Y}_a + p\hat{Y}_\ell$$

where $\hat{Y}_{nol}$ = the area frame estimate of the "nonoverlap" domain

$\hat{Y}_a$ = the area frame estimate of the "overlap" domain

$\hat{Y}_\ell$ = the list frame estimate of the "overlap" domain

q = weight attached to the area frame estimate of the "overlap" domain

p = weight attached to the list frame estimate of the "overlap" domain

$p+q=1$.

The variance of $\hat{Y}_H$ is:

$$(2)\quad Var(\hat{Y}_H) = Var(\hat{Y}_{nol}) + q^2 Var(\hat{Y}_a) + p^2 Var(\hat{Y}_\ell) + 2qCov(\hat{Y}_{nol},\hat{Y}_a)$$

where $Var(\cdot)$ denotes a variance and $Cov(\cdot,\cdot)$ denotes a covariance.

2. ALTERNATIVE EXPRESSIONS OF BASIC ESTIMATOR

One widely used multiple frame estimator is the "screening" estimator. The "screening" estimator is equation (1) with $p = 1$ and $q = 0$, i.e.:

$$\hat{Y}_{Screen} = \hat{Y}_{nol} + (0)\hat{Y}_a + (1)\hat{Y}_\ell = \hat{Y}_{nol} + \hat{Y}_\ell.$$

Obviously, the variance of $\hat{Y}_{Screen}$ is:

$$Var(\hat{Y}_{Screen}) = Var(\hat{Y}_{nol}) + Var(\hat{Y}_\ell)$$

because an area frame estimate is independent of a list frame estimate.

The area frame estimate of the total of the entire population may also be expressed in terms of equation (1) with $p = 0$ and $q = 1$, i.e.:

$$\hat{Y}_{Area} = \hat{Y}_{nol} + (1)\hat{Y}_a + (0)\hat{Y}_\ell = \hat{Y}_{nol} + \hat{Y}_a.$$

The variance of $\hat{Y}_{Area}$ is:

$$Var(\hat{Y}_{Area}) = Var(\hat{Y}_{nol}) + Var(\hat{Y}_a) + 2Cov(\hat{Y}_{nol},\hat{Y}_a).$$

Since $p+q=1$, an alternative expression for Hartley's estimator is:

$$\begin{aligned}\hat{Y}_H &= \hat{Y}_{nol} + (1-p)\hat{Y}_a + p\hat{Y}_\ell \\ &= \hat{Y}_{nol} + \hat{Y}_a + p(\hat{Y}_\ell - \hat{Y}_a) \\ &= \hat{Y}_{Area} + p(\hat{Y}_\ell - \hat{Y}_a)\end{aligned}$$

Similarly, the variance of $\hat{Y}_H$ may be written as:

$$(4)\quad Var(\hat{Y}_H) = Var(\hat{Y}_{Area}) - 2pCov(\hat{Y}_{Area},\hat{Y}_a) + p^2[Var(\hat{Y}_\ell) + Var(\hat{Y}_a)]$$

If one has the data to compute $\hat{Y}_a$, the "screening" estimator appears inefficient because it wastes this information. Better use may be made of data from both frames by combining $\hat{Y}_a$ and $\hat{Y}_\ell$ using an optimum p and q based on the minimization of the variance of $\hat{Y}_H$. From equation (4) optimum p is obtained as follows:

$$\frac{\partial Var(\hat{Y})}{\partial p} = -2pCov(\hat{Y}_{Area},\hat{Y}_a) + 2p[Var(\hat{Y}_\ell) + Var(\hat{Y}_a)]$$

Setting this equal to zero and solving for p gives:

$$(5)\quad p_{opt} = \frac{Cov(\hat{Y}_{Area},\hat{Y}_a)}{Var(\hat{Y}_a) + Var(\hat{Y}_\ell)}$$

or $$p_{opt} = \frac{Var(\hat{y}_a)+Cov(\hat{y}_{no\ell},\hat{y}_a)}{Var(\hat{y}_a)+Var(\hat{y}_\ell)}$$

Thus, Hartley's estimator, $\hat{y}_H$, is:

$$\hat{y}_H = \hat{y}_{Area} + p_{opt}\ (\hat{y}_\ell - \hat{y}_a)$$

and the variance in terms of p_{opt} is derived by substituting equation (5) into equation (4) so that:

$$(6)\quad Var(\hat{y}_H) = Var(\hat{y}_{Area}) - 2\,\frac{[Cov(\hat{y}_{Area},\hat{y}_a)]^2}{Var(\hat{y}_a)+Var(\hat{y}_\ell)} + \left[\frac{Cov(\hat{y}_{Area},\hat{y}_a)}{Var(\hat{y}_a)+Var(\hat{y}_\ell)}\right]^2 [Var(\hat{y}_a)+Var(\hat{y}_\ell)]$$

$$= Var(\hat{y}_{Area}) - \frac{[Cov(\hat{y}_{Area},\hat{y}_a)]^2}{[Var(\hat{y}_a)+Var(\hat{y}_\ell)]}$$

$$= Var(\hat{y}_{Area}) - p_{opt}\ Cov(\hat{y}_{Area},\hat{y}_a).$$

3. PROPOSED STRATUM MULTIPLE FRAME ESTIMATOR

The purpose of this paper is to extend the Hartley estimator to the case where the list frame is stratified. In this case:

$$\hat{y}_\ell = \hat{y}_{\ell(1)} + \hat{y}_{\ell(2)} + \ldots + \hat{y}_{\ell(k)}$$

where $\hat{y}_{\ell(h)}$ = list frame estimate of the total in the h^{th} stratum of the list, and there are k list strata. However, just as the area frame sample can be classified by domain to make an "overlap" estimate, $\hat{y}_a$, the area frame sample can also be matched with the corresponding list unit within the overlap domain to provide an "overlap" estimate for each stratum on the list:

$$\hat{y}_a = \hat{y}_{a(1)} + \hat{y}_{a(2)} + \ldots + \hat{y}_{a(k)}.$$

Using this list stratification yields an estimate of the form:

$$\hat{y}_{Strata} = \hat{y}_{no\ell} + \sum_{h=1}^{k} q_h\ \hat{y}_{a(h)} + \sum_{h=1}^{k} p_h\ \hat{y}_{\ell(h)},$$

where $q_h + p_h = 1$ for $h = 1, 2, \ldots, k$. The proposed estimator may also be written:

$$\hat{y}_{Strata} = \hat{y}_{no\ell} + \sum_{h=1}^{k} (1-p_h)\hat{y}_{a(h)} + \sum_{h=1}^{k} p_h\ \hat{y}_{\ell(h)}$$

$$= \hat{y}_{no\ell} + \sum_{h=1}^{k} \hat{y}_{a(h)} + \sum_{h=1}^{k} p_h\ [\hat{y}_{\ell(h)} - \hat{y}_{a(h)}]$$

$$= \hat{y}_{Area} + \sum_{h=1}^{k} p_h\ [\hat{y}_{\ell(h)} - \hat{y}_{a(h)}].$$

It is easier for the following work to change to matrix notation.

Let:

$$\underline{p} = \begin{bmatrix} p_1 \\ p_2 \\ \cdot \\ \cdot \\ \cdot \\ p_k \end{bmatrix}$$

$$\underline{\hat{y}}_\ell = \begin{bmatrix} \hat{y}_{\ell(1)} \\ \hat{y}_{\ell(2)} \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_{\ell(k)} \end{bmatrix} \qquad \underline{\hat{y}}_a = \begin{bmatrix} \hat{y}_{a(1)} \\ \hat{y}_{a(2)} \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_{a(k)} \end{bmatrix}$$

Then:

$$(7)\quad \hat{y}_{Strata} = \hat{y}_{Area} + \underline{p}'\ [\underline{\hat{y}}_\ell - \underline{\hat{y}}_a].$$

The variance of $\hat{y}_{Strata}$ is:

$$Var(\hat{y}_{Strata}) = Var(\hat{y}_{Area}) + \underline{p}'\ [\Sigma_\ell + \Sigma_a]\underline{p} + 2Cov(\hat{y}_{Area},\ \underline{p}'\ [\underline{\hat{y}}_\ell - \underline{\hat{y}}_a])$$

$$= Var(\hat{y}_{Area}) + \underline{p}'\ [\Sigma_\ell + \Sigma_a]\ \underline{p} - 2\underline{p}'\ Cov(\hat{y}_{Area},\underline{\hat{y}}_a)$$

where:

Σ_ℓ = variance-covariance matrix of $\underline{\hat{y}}_\ell$

$$\Sigma_\ell = \begin{bmatrix} Var(\hat{y}_{\ell(1)}) & 0 & \ldots & 0 \\ 0 & Var(\hat{y}_{\ell(2)}) & \ldots & 0 \\ \cdot & \cdot & \ddots & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & Var(\hat{y}_{\ell(k)}) \end{bmatrix}$$

Σ_a = variance-covariance matrix of $\hat{\underline{y}}_a$

$$\Sigma_a = \begin{bmatrix} Var(\hat{y}_{a(1)}) & Cov(\hat{y}_{a(1)},\hat{y}_{a(2)}) & \cdots & Cov(\hat{y}_{a(1)},\hat{y}_{a(k)}) \\ Cov(\hat{y}_{a(2)},\hat{y}_{a(1)}) & Var(\hat{y}_{a(2)}) & \cdots & Cov(\hat{y}_{a(2)},\hat{y}_{a(k)}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{y}_{a(k)},\hat{y}_{a(1)}) & Cov(\hat{y}_{a(k)},\hat{y}_{a(2)}) & \cdots & Var(\hat{y}_{a(k)}) \end{bmatrix}$$

and
$$Cov(\hat{y}_{Area},\hat{\underline{y}}_a) = \begin{bmatrix} Cov(\hat{y}_{Area},\hat{y}_{a(1)}) \\ Cov(\hat{y}_{Area},\hat{y}_{a(2)}) \\ \vdots \\ Cov(\hat{y}_{Area},\hat{y}_{a(k)}) \end{bmatrix}$$

We will require that Σ_ℓ and Σ_a be positive definite.

To find $\underline{p}$ we use the criterion of minimizing the variance of $\hat{y}_{Strata}$. Therefore, we set:

$$\frac{\partial Var(\hat{y}_{Strata})}{\partial \underline{p}} = \underline{0}$$

and solve for the optimum value of $\underline{p}$. Then

$$\frac{\partial Var(y_{Strata})}{\partial \underline{p}} = 0$$

yields:

$$2(\Sigma_\ell+\Sigma_a)\underline{p} - 2Cov(\hat{y}_{Area},\hat{\underline{y}}_a) = 0$$

$$(\Sigma_\ell+\Sigma_a)\underline{p} = Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

So, the optimum value of p is:

$$\underline{p}_{opt} = (\Sigma_\ell+\Sigma_a)^{-1}\ Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

Checking the second derivative verifies $\underline{p}_{opt}$ actually yields the minimum extremum of $Var(\hat{y}_{Strata})$.

4. COMPARING VARIANCES

Replace $\underline{p}$ with $\underline{p}_{opt} = (\Sigma_\ell+\Sigma_a)^{-1}Cov(\hat{y}_{Area},\hat{\underline{y}}_a)$ to yield:

(8) $$Var(\hat{y}_{Strata}) = Var(\hat{y}_{Area}) - Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'[\Sigma_\ell+\Sigma_a]^{-1}Cov(\hat{y}_{Area},\hat{y}_a),$$

or

$$Var(\hat{y}_{Strata}) = Var(\hat{y}_{Area}) - \underline{p}'Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

Since Σ_ℓ and Σ_a are positive definite, and thus $\Sigma_\ell + \Sigma_a$ and $(\Sigma_\ell+\Sigma_a)^{-1}$ are positive definite:

$$Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'[\Sigma_\ell+\Sigma_a]^{-1}Cov(\hat{y}_{Area},\hat{\underline{y}}_a)>0.$$

Therefore, equation (8) shows:

$$Var(\hat{y}_{Strata})<Var(\hat{y}_{Area}).$$

It is also possible to show that:

$$Var(\hat{y}_{Strata})\leq Var(\hat{y}_H).$$

Remember:

$$Var(\hat{y}_H)=Var(\hat{y}_{Area}) - \frac{[Cov(\hat{y}_{Area},\hat{y}_a)]^2}{[Var(\hat{y}_a)+Var(\hat{y}_\ell)]}$$

and:

$$Var(\hat{y}_S) = Var(\hat{y}_{Area})-Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'[\Sigma_a+\Sigma_\ell]^{-1} Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

Therefore, $Var(\hat{y}_{Strata}) \leq Var(\hat{y}_H)$ if:

(9) $$\frac{[Cov(\hat{y}_{Area},\hat{y}_a)]^2}{[Var(\hat{y}_a)+Var(\hat{y}_\ell)]} \leq Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'[\Sigma_\ell+\Sigma_a]^{-1}\ Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

A change of notation for Hartley's estimator will make this proof clear.

$$Cov(\hat{y}_{Area},\hat{y}_a)=Cov(\hat{y}_{Area},\hat{y}_{a(1)}+\hat{y}_{a(2)}+\ldots+\hat{y}_{a(k)})$$
$$=Cov(\hat{y}_{Area},\hat{y}_{a(1)})+Cov(\hat{y}_{Area},\hat{y}_{a(2)}) + \ldots + Cov(\hat{y}_{Area},\hat{y}_{a(k)})$$

$$=Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'\underline{1}.$$

and,

$$Var(\hat{y}_a)+Var(\hat{y}_\ell) = \sum_{h=1}^{k}\sum_{h'=1}^{k} Cov(\hat{y}_{a(h)},\hat{y}_{a(h')})$$

$$+\sum_{h=1}^{k} Var(\hat{y}_{\ell(h)})$$

$$= \underline{1}'\Sigma_a\underline{1} + \underline{1}'\Sigma_\ell\underline{1}$$

$$= \underline{1}'(\Sigma_a+\Sigma_\ell)\underline{1}.$$

Thus, proving equation (9) resolves into proving:

$$(10)\quad \frac{[Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'\underline{1}]^2}{\underline{1}'(\Sigma_a+\Sigma_\ell)\underline{1}} \leq Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'(\Sigma_a+\Sigma_\ell)^{-1}Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

A theorem from matrix theory (Rao, page 60) says that if A is a positive definite mxm matrix, and $\underline{u}$ and $\underline{x}$ are m-vectors, then:

$$(11)\quad \frac{(\underline{u}'\underline{x})^2}{\underline{x}'A\underline{x}} \leq \underline{u}'A^{-1}\underline{u}.$$

This theorem is a result of the Cauchy-Schwartz inequality. Substituting

$$A = \Sigma_a + \Sigma_\ell,$$

$$\underline{u} = Cov(\hat{y}_{Area},\hat{\underline{y}}_a)$$

$$\underline{x} = \underline{1},$$

into inequality (11) gives

$$\frac{[Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'\underline{1}]^2}{\underline{1}'[\Sigma_a+\Sigma_\ell]\underline{1}} \leq Cov(\hat{y}_{Area},\hat{\underline{y}}_a)'(\Sigma_a+\Sigma_\ell)^{-1}Cov(\hat{y}_{Area},\hat{\underline{y}}_a).$$

which is exactly inequality (10) completing the proof that:

$$Var(\hat{y}_{Strata}) \leq Var(\hat{y}_H).$$

5. EMPIRICAL STUDY

To illustrate how the four estimators $\hat{y}_{Area}$, $\hat{y}_{Screen}$, $\hat{y}_H$, and $\hat{y}_{Strata}$ compare, data for cattle and hogs from June 1974 area and list frame surveys conducted by the Statistical Reporting Service, U.S.D.A., in a midwestern state, were used to obtain totals and variances for each estimator. The estimates and standard errors from each frame estimating the overlap domain strata are presented in Table 1. These estimates are substituted into the $\hat{\underline{y}}_\ell$ and $\hat{\underline{y}}_{o\ell}$ vectors for the multiple frame estimator $\hat{y}_{Strata}$. Estimates and standard errors vary considerably between frames for the same stratum. Standard errors are much smaller for the list frame estimates in the larger strata than for the area frame estimates.

The variances of $\hat{y}_H$ and $\hat{y}_{Strata}$ are determined by $Var(\hat{y}_{Area}) - p_{opt}[Cov(\hat{y}_{Area},\hat{y}_a)]$ and $Var(\hat{y}_{Area}) - \underline{p}'_{opt}[Cov(\hat{y}_{Area},\hat{\underline{y}}_a)]$ respectively. The p_{opt} values, correlations between segment totals and overlap domain, and the amounts by which $Var(\hat{y}_{Area})$ is reduced to equal $Var(\hat{y}_H)$ and $Var(\hat{y}_{Strata})$ are presented for the estimators $\hat{y}_{Strata}$ and $\hat{y}_H$ in Table 2. Values of p_{opt} differ considerably between strata for the $\hat{y}_S$ estimator and from the p_{opt} obtained for $\hat{y}_H$. The value of $Cov(\hat{y}_{Area},\hat{y}_a)$ summed over the strata is the same as for $\hat{y}_H$. Therefore, the p_{opt} values optimized on a stratum-by-stratum basis are weighted by the $Cov(\hat{y}_{Area},\hat{y}_a)$ values for each stratum. The larger the weighted value of p_{opt} for $\hat{y}_{Strata}$ compared to the unweighted p_{opt} of $\hat{y}_H$ the smaller the variance of $\hat{y}_{Strata}$ relative to $Var(\hat{y}_H)$.

TABLE 1--List and Area Livestock Inventory Estimates by Livestock Stratum, June 1974

	Overlap Domain					
	List ($\hat{y}_{\ell}$)			Area ($\hat{y}_a$)		
Multiple Frame Strata	Stratum Estimates	Standard Error	Coefficient of Variation	Stratum Estimates	Standard Error	Coefficient of Variation
	(000)	(000)	(%)	(000)	(000)	(%)
Hogs & Pigs						
1 (Unknown)	44.0	8.5	19.3	44.7	22.8	51.1
2 (0-9 Hogs)	305.4	70.0	22.9	200.3	50.5	25.2
3 (10-49 Hogs)	140.1	26.9	19.2	195.1	76.6	40.8
4 (50-499 Hogs)	314.4	24.3	7.7	328.9	110.8	33.7
Cattle & Calves						
1 (Unknown)	363.6	56.5	15.5	175.8	33.2	18.9
2 (0-9 Cattle)	436.3	69.4	15.9	327.3	60.2	18.4
3 (10-49 Cattle)	1179.8	53.1	4.5	975.2	111.1	11.4
4 (50-499 Cattle)	1334.8	56.7	4.2	1263.7	180.9	14.3

TABLE 2--Multiple Frame Components for Estimates $\hat{y}_H$ and $\hat{y}_S$, June 1974

Estimator	Stratum 1/	p_{opt} 2/	$Corr(\hat{y}_{Area}, \hat{y}_a)$ 3/	$p_{opt}[Cov(\hat{y}_{Area}, \hat{y}_a)]$ 4/	Percent of $Var(\hat{y}_{Area})$ 5/
				(000,000)	(%)
	1	1.216	.187	918	3
	2	.324	.300	862	3
Hogs	3	.989	.583	7,806	25
	4	.995	.689	13,415	42
$\hat{y}_S$	Total	.927 6/	-	23,001	73
$\hat{y}_H$	Overall	.829 7/	.913	20,556	65
	1	.288	.152	337	1
	2	.406	.201	1,139	2
Cattle	3	.806	.388	8,034	15
	4	.915	.719	27,517	51
$\hat{y}_S$	Total	.841 6/	-	37,027	69
$\hat{y}_H$	Overall	.774 7/	.920	34,077	64

1/ List frame strata based on number of head in the operation.
2/ Weight attached to the list frame estimate.
3/ Correlation between area frame total and area frame overlap domain.
4/ Amount by which $Var(\hat{y}_{Area})$ is reduced to equal $Var(\hat{y}_{Strata})$ and $Var(\hat{y}_H)$.
5/ p_{opt} $[Cov(\hat{y}_{Area}, \hat{y}_a)]$ as percent of $Var(\hat{y}_{Area})$.
6/ Weighted value of p_{opt} for $\hat{y}_{Strata}$ from weighing individual stratum p_{opt} by stratum $Cov(\hat{y}_{Area}, \hat{y}_a)$.
7/ Value of p_{opt} for $\hat{y}_H$ ignoring strata.

Individual stratum p_{opt} values greater than 1 in effect give a negative weight to the area frame as in Stratum 1 for hogs and pigs. The size of p_{opt} depends upon $Var(\hat{y}_{o\ell})+Cov(\hat{y}_{no\ell},\hat{y}_{o\ell})$ relative to $Var(\hat{y}_{\ell})+Var(\hat{y}_{o\ell})$. If the covariance between the nonoverlap domain and the overlap domain for a given stratum is larger than the variance of the list estimate for that stratum then p_{opt} is greater than one. It is also noteworthy that for both hogs and cattle the contribution of the small livestock strata in reducing the total variance is minimal. This was due to smaller overlap domain standard errors for these strata relative to the higher strata and poor correlation between $\hat{y}_{Area}$ and $\hat{y}_{a}$ in the lower strata.

Except for Stratum 1 Hogs, the weight for the list frame, p_{opt}, and the correlation, $Corr(\hat{y}_{Area},\hat{y}_{a})$, were lower for the smaller strata. Even when the list frame weight was greater than one, the reduction in variance was small. The stratum-by-stratum multiple frame procedure provides a means of measuring the contribution of the frames for each stratum.

The combinations of estimates from each of the two frames into the various multiple frame estimates are presented in Table 3. There is little difference between results for the screening estimator $(\hat{y}_{Screen})$ and the Hartley estimator $(\hat{y}_{H})$. The weight attached to the list frame is so dominant that very little reduction in sampling error is realized from the contribution of the area frame overlap domain.

TABLE 3--Multiple Frame Livestock Estimates Using Alternative Estimators, June 1974

Multiple Frame Estimator	Hogs			Cattle		
	Estimate	Standard Error	Coefficient of Variation	Estimate	Standard Error	Coefficient of Variation
	(000)	(000)	(%)	(000)	(000)	(%)
$\hat{y}_{Area}$	1301.1	177.4	13.6	3615.4	231.4	6.4
$\hat{y}_{Screen}$	1299.9	106.6	8.4	4188.9	149.6	3.6
$\hat{y}_{H}$	1300.1	104.4	8.0	4067.1	139.5	3.4
$\hat{y}_{Strata}$	1265.4	92.0	7.3	3952.2	128.6	3.3

The $\hat{y}_{Strata}$ estimate for hogs was slightly below the other multiple frame estimates while the $\hat{y}_{Strata}$ cattle estimate was between $\hat{y}_{Area}$ and $\hat{y}_{Screen}$, and near $\hat{y}_{H}$. This reflects the relative size of the area frame estimate and its weight compared to the list frame estimate and its weight in each stratum.

The stratum-by-stratum combination of area and list frame estimates resulted in the smallest sampling errors for $\hat{y}_{Strata}$ as expected. Sampling errors on hog and cattle estimates were about 14% lower for $\hat{y}_{Strata}$ than the screening estimator, $\hat{y}_{Screen}$, and approximately 12% and 8% respectively below $\hat{y}_{H}$.

REFERENCES

[1] Hartley, H. O., (1962) "Multiple Frame Surveys," Proceedings of the Social Science Section of the American Statistical Association Meeting, Minneapolis, Minnesota.

[2] Rao, C. R., (1973) Linear Statistical Inference and Its Applications, John Wiley & Sons, New York, N.Y.

[3] Fuller, Wayne A. and Burmeister, Leon F., (1973) "Estimates for Samples Selected From Two Overlapping Frames," A report for the U.S. Bureau of the Census, Iowa State University, Ames, Iowa.

[4] Cochran, Robert S., (1964) "Multiple Frame Sample Surveys," Proceedings of Social Science Section of American Statistical Association Meeting, Chicago, Illinois.

METHODS OF ESTIMATING INCIDENCE OF BURN INJURIES

D. M. Branson, P. H. Darling and M. H. Branson
University of Rhode Island

I. INTRODUCTION

In recent years, the serious problem of burn injuries and deaths has been frequently debated among a wide spectrum of special interest groups. Estimates of the number of injuries and deaths attributable to fire and flame accidents in general, and to flammable fabrics in particular are often quoted, yet rarely related back to a specific population.

The 1953 Flammable Fabrics Act was expanded in 1967 with the Secretary of Commerce being given the authority to act in the public interest if he finds there is unreasonable risk to the public arising from the use of one or more flammable fabrics. Since 1967, several flammability standards have been promulgated including two apparel standards requiring children's sleepwear, sizes 0-14, to be flame-retardant. The children's sleepwear standards have meant increased costs for the consumer due to a higher initial price, reduced wear-life and a reduction in choice. The benefits of such legislation include decreased incidence of burn cases and/or decreased severity of injuries. Cost-benefit analysis could be used in determining whether such programs are in the consumer's interest.

The major thrust of this research was directed toward examining how well incidence could be measured from existing data sources. It should be noted that good estimates of incidence rates represent only one input necessary in a detailed cost-benefit analysis. A number of other studies (3, 11) have addressed the problem. The Pittsburgh Burn Study (11) represents an example of a study undertaken in a specific geographical and time setting.

The Consumer Product Safety Commission and the Department of Commerce are operating an ongoing data collection system for the purpose of gathering information on product related injuries. The establishment of the National Electronic Injury Surveillance System (NEISS), a computerized data-gathering system operative in 119 statistically selected emergency rooms in the United States, should aid in the acquisition of reliable data (10). However, a limitation of the system for determining incidence of burn injuries is that a burn victim entering the hospital by direct admission or through a special burn unit would not be reported by N E I S S. Therefore, the NEISS data yields a lower bound estimate of incidence of burn victims.

Dardis and Schmitt have examined the role of cost-benefit analysis in evaluating flammability standards (3). Their results were designed to illustrate methodology rather than to determine whether the 0 to 6x children's Sleepwear Standard was justified.

Research Setting

Rhode Island is entirely divided into census tracts with census tract recorded in most routinely collected health data. The characteristics of the people closely approximate those of the nation as a whole (2).

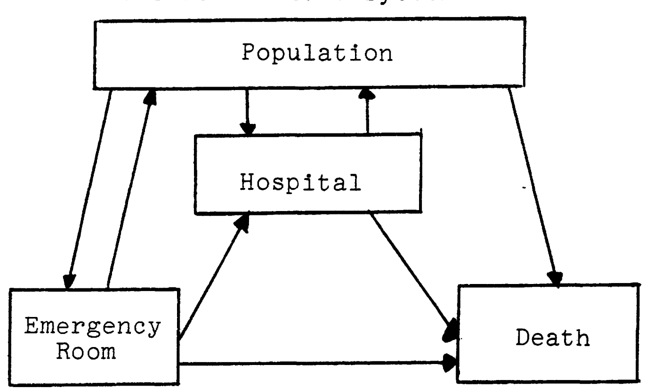
Substantial health data is routinely collected and available through the statewide Center for Health Statistics at Rhode Island Health Services Research, Inc. (SEARCH). In addition, the health care system is a closed system in that most residents receive their medical care within the state. The volume of medical care delivered to residents in nearby states has been documented in other studies.

Flow Diagram for Burn Injuries

In measuring incidence, the flow patterns of burn victims into and through the health care system must be considered. A simplified representation is given in Figure 1.

FIGURE 1

Flow of Burn Injuries in the Health Care System



Burn injuries may first enter the health care system via the emergency room or by direct hospital admission. A burn injury arriving at the emergency room may be cared for and released, admitted to the hospital, or be dead on arrival. A direct hospital admission may be treated for a period of time and released or the victim may die while hospitalized.

Clearly then, to get an accurate estimate of incidence, each of these flows must be carefully tracked. Victims appearing in more than one segment of the system must be separated out. Otherwise some cases will be counted more than once. Also, the estimate of incidence must be related back to the population at risk. For this study, all Rhode Island residents represented the population at risk.

II. EXISTING DATA SOURCES

Census Data

The 1970 census data, tabulated by census tracts were utilized as the basis for the population at risk. However, some Rhode Island hospitals provided medical services to nearby Connecticut and Massachusetts residents. And conversely some RI residents utilized out-of-state hospitals for their medical care. In an effort to rectify these two problems, one of which would affect the number of burn victims and the other which would affect the number in the "population at risk," two methods were employed. First, to correct for RI residents receiving out-of-state medical care - the population in each census tract was adjusted downward to reflect the proportion of each tract's population that sought its medical care outside the state. The adjustment factors for each city and town (5), were applied uniformly to each census tract within the city or town and to each age group within the tract. To correct for out-of-state residents who received care in RI hospitals, these victims were not included in the calculations of the numerators for incidence rates.

In a study, sponsored jointly by the Rhode Island Department of Health and SEARCH, the total Rhode Island population was divided into four socio-economic groups by census tracts. Using factor analysis, Sakoda and Karon (7) developed a socio-economic status factor score for each census tract. The scores were aggregated into four groupings. The population was thus divided as follows: high - 25.6 percent of the population, middle - 38.6 percent, low 28 percent and poverty 7.8 percent, a percentage approximating the number of people on welfare (8). The socio-economic status characteristics of the RI burn population could therefore be studied and analyzed by knowing the patient's census tract.

Vital Statistics

Data on victims who died as a result of their burn injuries for the years 1972 through 1974 were studied.

Hospital PAS Data

Hospital PAS data (4) provided invaluable data on patients hospitalized with a primary diagnosis of burn injury in Rhode Island for the years 1972 through 1974. This was especially helpful as all RI general acute care hospitals are participating PAS hospitals. Most of the desired variables for each hospitalized burn case were available in computer format. However, additional variables were deemed important to the study as it was desirable to examine the role that clothing played in the burn incident. Previous studies (1, 9, 11) had indicated that if clothing ignition occurred, a more severe burn resulted than if the clothing had not ignited. Source of the burn injury and severity of the injury were two other variables worthy of exploration.

In addition, incidence rates for a population based on PAS discharge records alone are subject to two main deficiencies. First, it was difficult to discern the number of "distinct" victims that the PAS abstracts represented; this problem stemmed from victims being readmitted one or more times, with a separate PAS form being completed after each discharge. Second, it was difficult to determine the number of "new" burn cases; the PAS classification system allowed for the coding of a late effect plus an additional code, such as for reconstructive surgery. However the hospitals did not uniformly follow the recommended procedure.

Therefore, one Rhode Island hospital was contacted for a preliminary investigation of their medical records to determine what data were available, in what manner it was available and how difficult it would be to retrieve the additional variables. Through this investigation it was deemed feasible to abstract a sample of individual medical records from the burn population to determine: 1. clothing involvement, 2. source of the burn, 3. type of admission, 4. admission number and 5. severity of burn. Results based on analysis of clothing involvement and source of the burn are given elsewhere (2).

Hospital Medical Records

There were fifteen acute care hospitals in Rhode Island. Hospitals with very few burn injuries or with numerous out-of-state patients were not contacted. Each hospital contact was handled in an individual manner. Because most of the physicians contacted at the hospitals were pediatricians, they advised that the study be limited to the age range 0-19.

Therefore, the sample consisted of victims with a primary diagnosis of burn injury, ages 0 to 19, for the years 1972 through 1974.

Permission was granted by all sever contacted hospitals with the understanding that patient confidentiality would be maintained. These seven hospitals accounted for approximately 79% of the hospitalized Rhode Island burn victims in the 0 to 19 age range.

III. ESTIMATION OF BURN INCIDENCE RATE FOR AGES 0-19

Using the sampled data from medical records, certain estimates of victims within components of the health care delivery system were derived. These estimates were based on the assumption that the results of the sample (approximately 79 percent of the hospitalized Rhode Island burn patients, ages 0-19) were representative of the total number of youths, aged 0-19, who were hospitalized with a burn injury. However, the sample was not a random sample.

A summary of data from the sampled burn population is found in Table 1. The sample data were first segregated into first hospital episodes and subsequent episodes. First hospitalizations were then divided into those admitted through the emergency room and those admitted directly. Percentages were calculated for each group and then applied to the total number of PAS discharges to obtain the estimates given in Table 2. For example, to obtain an estimate of the total number of first admissions, 346 was multiplied by 91.5 percent, the percent of first admissions determined from the sample.

These estimates were combined with data from the other three existing data sources to approximate the number of victims in the various segments of the health care system, as seen in Figure 2. The numbers found in the sample could not be used directly since the sample contained only 79 percent of the RI burn population in the 0 to 19 age group.

Twenty-one Rhode Islanders, age 0 to 19, died from burn injuries. Vital Statistics indicated that ten of these deaths occurred in hospitals, ten deaths occurred at home and one youth was dead on arrival at a hospital. PAS data reported five hospital deaths, providing the number of victims who died as a result of their injuries while being cared for in the hospital. This number did not include victims who died in the emergency room, as PAS was designed for use with inpatient hospitalizations. Therefore, it was assumed that the five remaining hospital deaths, as reported by Vital Statistics, occurred in the emergency room. This assumption was considered reasonable as hospital discharge status and location of death were generally considered to be reliable data items. These results are summarized in Table 3.

TABLE 1

Sample Hospital Data
Ages 0-19

Total Sampled Rhode Island Residents	272
First Admissions Through the Emergency Room	207
First Admissions Admitted Directly	42
Total First Admissions	249
Percentage of First Admissions	91.5%
Percentage of Subsequent Admissions	8.5%
Percentage of First Admissions Through the Emergency Room	83.1%
Percentage of First Admissions Admitted Directly	16.9%

TABLE 2

Total Hospital Data and Applied
Percentages, Ages 0-19

Total of PAS Hospitalized Cases	346
Estimate of First Admissions	317
Estimate of First Admissions Through the Emergency Room	264
Estimate of First Admissions Admitted Directly	53

To estimate the total number of burn injuries, the 317 hospitalized cases were added to the sixteen victims who died prior to hospital admittance for a total of 333 victims, or 111 victims per year. Census data, adjusted downward as previously described, yielded an estimate of 324,838 Rhode Island youth, aged 0 to 19. This number represented the population at risk.

TABLE 3

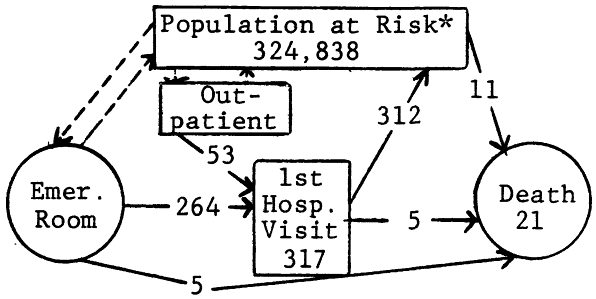
Death Data, Ages 0-19

Total Number of Deaths	21
Number Died at Home	10
Number Dead on Arrival	1
Number Died in the Hospital (V.S.)	10
Number Died in the Hospital (PAS)	5
Estimated Number Died in the Emergency Room	5

The dotted lines in Figure 2 indicate information that was not collected. The current research was not able to estimate the incidence of burn victims cared for in the emergency room, or some other outpatient setting, and released. Nevertheless, information derived through the

FIGURE 2

Flow of Burn Victims
(RI Residents, Ages 0-19)
1972-1974



*All estimates, excluding the population at risk, represent three year totals.

sample yielded 42 hospitalized cases that had been seen in physician's office, clinic, etc. for initial care after the accident. These victims were later admitted to the hospital directly, due to complications. Obviously, many more injuries were treated in outpatient settings that did not later require hospitalization. The present research concentrated only on burn injuries sufficiently serious to warrant hospitalization.

It should be noted that 16 of the 21 victims who died as a result of their injuries, died before being admitted to the hospital. Of the sampled burn population, 16.9 percent were admitted to the hospital directly. If this percentage were representative of the burn population in all age categories in all parts of the country, then the NEISS data collection system would be underestimating the problem of burn injuries by almost 17 percent. However, such a broad generalization can not be made, since the practices of the people seeking medical expertise and the physicians delivering medical care are not uniform throughout the United States.

The incidence rate of Rhode Island youths, ages 0 to 19, who suffered a burn injury resulting in hospitalization or death, was estimated to be 34.2 per 100,000. This estimated rate represents an average for a three year perioa, 1972 through 1974.

IV SELECTED DEMOGRAPHIC CHARACTERISTICS OF BURN VICTIMS

In the previous section, the incidence rate for burn injuries and deaths in the 0 to 19 year age group were presented. This required the combining of data from four sources including the sample taken from medical records at selected hospitals. If one utilizes only the first three data sources, demographic characteristics of all burn victims can be studied. It should be noted that the rates presented in this section do not represent incidence rates in all cases as hospitalization rates include subsequent re-admissions for some burn victims. However, with this in mind, useful insights can be gained using the readily accessible data alone.

Age and Sex Distribution Of Burn Deaths

Table 4 presents the death rates for Rhode Island residents by age and sex. There were a total of 86 deaths attributable to burn injuries for the three years studied. The death rates calculated, represent three year average rates. The elderly suffered a higher death rate than the other age groups. This finding further substantiated the vulnerability of the aged. The death rates for males exceeded the death rate for females in all age categories but the 5 to 14 year olds. It should be noted however that these rates are based on a small number of events but represent the actual population experience. Due to the small numbers of events, no further breakdown of data on death victims was attempted.

TABLE 4

Age/Sex Specific Death Rates for RI Residents

Sex	Age						Sex Specific
	0-4	5-14	15-19	20-44	45-64	65+	Death Rates
Male	4.4	.4	3.9	3.1	4.9	11.6	3.9
Female	1.9	2.8	.8	1.0	3.1	4.9	2.4

Age and Sex Distribution of Hospital Cases for 1972-1974

Of the 807 patients discharged from Rhode Island hospitals during the years of 1972 through 1974, with a primary diagnosis of burn injury, 738 were Rhode Island residents at the time of their hospitalizations. Using these cases, an estimate of the age/sex specific incidence rates of burn injuries was derived for each year studied (Table 5). It should be recalled that from the sampled hospital data, it was learned that 8.5 percent of the total number of cases sampled were not hospitalizations for the first episode of the burn injury. No determination of subsequent hospitalizations for other age groups was made, as the sample included only patients age 0-19. Therefore, the rates given in Table 5 were calculated using the total number of PAS discharges in each age/sex category. Although it was recognized that incidence would thus be over-estimated, nevertheless it was decided that to include all hospital discharges for some age groups and only first hospital episodes for other age groups would be biasing the results.

TABLE 5
Age/Sex Specific Rates For Burn Hospitalizations by Year for RI Residents (Per 100,000)

	0-4	5-14	15-19	20-44	45-64	65+	Sex Specific	Total
1972								
Males	93.2	28.0	27.9	28.2	25.4	32.4	33.4	
Females	80.9	14.6	9.9	12.8	8.4	13.1	17.1	25.1
1973								
Males	85.2	22.2	37.2	35.6	26.4	20.0	34.0	
Females	66.9	8.5	9.9	7.9	14.9	16.4	15.4	24.5
1974								
Males	90.5	36.1	41.9	45.7	34.9	37.4	44.2	
Females	75.3	9.7	29.8	11.4	8.4	19.6	18.0	30.9

In both sexes, the incidence rate for children four years of age and younger was at least 2.5 times greater than any other age group. The incidence rate for other age groups was fairly constant for males and females respectively.

Further examination of Table 5 indicates that the age group 0 to 4 was the only age group in which the incidence rates for males and females were similar. In all other age groups, the male incidence rate was at least twice the female incidence rate. Why males experienced burn injuries more frequently than females could not be satisfactorily explained and remains an area for further study.

TABLE 6
Age/Sex Specific Rates for Burn Hospitalizations by Socio-Economic Status Groups (three year average 1972-1974)

	0-4	5-14	15-19	20-44	45-64	65+	Total by Sex*	Total **
High								
Male	71.1	15.8	21.4	21.5	19.3	23.1	24.3	19.3
Female	53.0	6.0	19.2	13.0	5.1	20.3	14.5	
Middle								
Male	42.6	29.4	34.2	28.3	31.0	26.9	30.7	22.6
Female	61.5	9.1	13.2	9.3	13.6	10.5	14.7	
Low								
Male	116.1	29.9	50.2	38.6	27.5	33.2	41.7	30.7
Female	89.2	17.0	13.2	9.8	7.7	24.5	18.9	
Poverty								
Male	260.4	70.2	50.3	131.4	54.1	44.5	98.8	61.9
Female	143.1	17.1	33.4	12.8	22.1	5.1	26.5	

*Sex Specific, age adjusted incidence rates.
**Age/Sex adjusted incidence rates.

Socio-Economic Status Characteristics of the Burn Population

In the available hospital data, there was no direct measure of an individual's socio-economic status; such as income, occupation, education or net worth. Using the patients' census tract which was available as part of the PAS data, the 738 RI hospitalized burn victims, were placed in one of the four SES groups as was described earlier.

Using the 1970 census data adjusted for RI residents going out of state, age/sex specific incidence rates by socio-economic status were calculated per 100,000. Yearly rates were not derived due to the small number of individuals within some cells.

By examining Table 6, a definite trend becomes apparent. The high SES group had an overall rate of 19.3, the middle 22.6, the low 30.7 and the poverty group an alarming 61.9.

The high incidence rates for poverty males should be noted. The age group 0 to 4 and 20 to 44 experienced particularly high hospitalization rates. The causes behind the rates should be further examined.

V CONCLUSIONS

Good estimates of incidence rates require good estimates of the numerator (the number of new burn cases occurring during the specified time period) and the denominator (the number of Rhode Island residents that could have sustained a burn injury during the specified time period). Wherever possible, existing data sources should be utilized.

Hospital PAS data, provided invaluable data on patients hospitalized with a primary diagnosis of burn injury in Rhode Island. This was especially so since all Rhode Island general acute care hospitals are participating PAS hospitals.

A second source of hospital data was needed to compensate for some of the deficiencies of PAS data. This required abstracting data from medical records, a time consuming process which is not feasible on an ongoing basis. In addition, it is necessary to obtain permission from individual hospitals before accessing the records.

Mortality and census data used in this study were readily available in suitable form. Only minor problems were encountered with these sources. The census data was for 1970 and as the RI population is fairly stable, it could be used with reasonable confidence in calculating rates.

As a result of the shortcomings of the existing data sources, it is difficult to obtain an accurate estimate of the incidence rate of burn injuries in the given population. It is necessary to supplement existing data with a special study as was done in this work. However, with this additional data, more detailed information was available.

Periodic studies could be conducted to augment existing data and to determine if the percentages derived from this work are applicable to other age groups and geographical localities. This approach has the potential for offering reasonably good estimates of incidence.

REFERENCES

1. America Burning. The Final Report of the National Commission on Fire Prevention and Control. Quoted in Burn Prevention Fact Sheet. Boston: Shriners Burn Institute, May, 1973, pp. 1-4.

2. Branson, Donna M., "Estimating Population Based Incidence Rates for Burn Cases in Rhode Island," Master's Thesis, University of Rhode Island, 1976.

3. Dardis, R. and Schmitt, J., "Cost Benefit Analysis of Flammability Standards." Proceedings of American Association of Textile Chemists and Colorists Meeting. Chicago, October, 1975.

4. PAS Professional Activity Study, MAP Medical Audit Program. Ann Arbor: Commission on Professional and Hospital Activities, 1971.

5. Patient Origins of General Hospital Admissions...Rhode Island. Rhode Island Department of Health, 1971.

6. Rhode Island - The Picture of Health. Rhode Island Department of Health, Office of Comprehensive Health Planning (January, 1975).

7. Sakoda, James, and Karon, James. Factor Analysis of Rhode Island Census Tracts, 1970. Office of Comprehensive Health Planning of the Rhode Island Department of Health, Rhode Island Health Services Research, Inc. (March, 1974).

8. Scott, Denman and MacKie, Anita, "Decisions to Hospitalize and Operate: A Socio-Economic Perspective in an Urban State." Surgery, Vol. 77, No. 2 (February, 1975), pp. 311-317.

9. Textile Industry Product Safety Committee. Fabric Flammability: What the Consumer Should Know. Washington, D.C. 1973.

10. U.S. Consumer Product Safety Commission Publication, "NEISS: An Overview." NEISS News, Vol. 3, No. 5 (April, 1975).

11. U.S. Department of Health, Education and Welfare. Fourth Annual Report to the President and the Congress on the Studies of Deaths, Injuries, and Economic Losses Resulting from Accidental Burning of Products, Fabrics or Related Materials, 1972.

R. L. Carter, Florida University
W. A. Fuller, Iowa State University

ABSTRACT

Alternative instrumental variable estimators for the slope in the simple errors in variables model are discussed. The modified limited information estimator is used to construct a randomly weighted average estimator similar to those studied by Huntsberger. The maximum likelihood estimator is derived for the case in which the error covariance is known to be zero. The limiting distributions of these estimators are obtained. The modified maximum likelihood estimators, the randomly weighted average estimators and Feldstein's randomly weighted average estimator are compared in a Monte Carlo study.

1. INTRODUCTION

The errors in variables model differs from the classical linear regression model in that the independent variables are not observed directly, but are masked by measurement error. For such models it is well known that the ordinary least squares estimators are biased. Generally speaking, estimation of the parameters of such a model requires additional information. This information may take the form of knowledge of the error variances, observations on other variables which are uncorrelated with the measurement errors, or knowledge of the form of the distribution of the errors and (or) regression variables. The case in which observations on additional variables, called instrumental variables, are available is considered in this paper. For discussions of the other two cases see [2], [6], [12] and [13].

The use of instrumental variables to estimate the parameters of errors in variables models has been heavily used in economics. Reiersol [14], Geary [7], Durbin [3] and Sargan [15] have discussed this method and most modern econometric texts contain a discussion of instrumental variable estimation. See, for example, Johnston [11] and Goldberger [8].

The methods used to estimate a single equation in a system of equations are closely related to the method of instrumental variables. In particular, for a just identified equation the two stage least squares and limited information maximum likelihood procedures reduce to the instrumental variable procedure. It is well known that these procedures yield estimators that may not possess moments in finite samples. Fuller [5] gave a modification of the limited information estimator and derived its mean square error through terms of order T^{-2}.

Huntsberger [10] investigated estimators of the mean of the normal population, constructed as randomly weighted averages of two estimators. One estimator, denoted by $\hat{\theta}_1$, is unbiased for the mean, θ_1, and has variance A_1. The other estimator, $\hat{\theta}_2$, is independent of $\hat{\theta}_1$ and has an unknown mean, θ_2, and variance A_2. We denote Huntsberger's family of estimators by

$$\hat{\theta}_H = \hat{a}\,\hat{\theta}_1 + (1-\hat{a})\hat{\theta}_2, \tag{1.1}$$

where the weight, $\hat{a}$ is a function of the sample. The weight

$$\hat{a}_w = \frac{(\hat{\theta}_1 - \hat{\theta}_2)^2 + A_2}{A_1 + A_2 + (\hat{\theta}_1 - \hat{\theta}_2)^2} \tag{1.2}$$

is an estimator of the fixed weight that minimizes the mean square error of a fixed weight estimator of the form (1.1). We call an estimator of this type a randomly weighted estimator. Huntsberger demonstrated that the mean square error of $\hat{\theta}_H$ is a function of γ, where

$$\gamma = (A_1 + A_2)^{-\frac{1}{2}} \, |\theta_1 - \theta_2| \, . \tag{1.3}$$

The mean square error of $\hat{\theta}_H$ is less than that of $\hat{\theta}_1$ if $\gamma=0$. As γ increases the mean square error rises to a maximum above the mean square error of $\hat{\theta}_1$. If $A_1=A_2$ and the estimator with $\hat{a}_w$ is used the maximum occurs near $\gamma=2.75$. As γ increases the mean square error of $\hat{\theta}_H$ approaches that of $\hat{\theta}_1$. Changing the form of $\hat{a}$ alters the mean square error function somewhat, but in all cases the function has the distinctive shape described above.

Feldstein [4] recently proposed an estimator for the measurement error model with an instrumental variable present. The estimator is an average of the instrumental variable and ordinary least squares estimators similar to (1.1). Therefore, one might expect the behavior of the estimator to depend in a critical manner on the true parameters. Feldstein, however, presented a Monte Carlo study showing his estimator to be uniformly superior to the instrumental variable estimator.

For the errors in variables model with one independent variable we compare Fuller's estimator with randomly weighted average estimators of the type considered by Huntsberger and Feldstein. We shall identify two models which result in slightly different estimation procedures.

2. THE MODELS

Let the errors in variables model be defined by

$$\underset{\sim}{Y} = \underset{\sim}{x}\beta + \underset{\sim}{e}\, , \qquad \underset{\sim}{X} = \underset{\sim}{x} + \underset{\sim}{u}\, , \tag{2.1}$$

where

- $\underset{\sim}{Y}$ is a Tx1 vector of observations on the dependent variable Y,
- $\underset{\sim}{x}$ is a Tx1 vector of unobservable true values of the independent variable x,
- $\underset{\sim}{e}$ is a Tx1 vector of errors,
- $\underset{\sim}{X}$ is a Tx1 vector of observations on the observable random variable X,
- $\underset{\sim}{u}$ is a Tx1 vector of measurement errors.

Rewriting the model in terms of observable random variables we have:

$$\underset{\sim}{Y} = \underset{\sim}{X}\beta + \underset{\sim}{w}, \text{ where } \underset{\sim}{w} = \underset{\sim}{e} - \underset{\sim}{u}\beta \, .$$

We assume the existence of a Tx1 vector of observations on an instrumental variable, $\underset{\sim}{z}$, with t^{th} observation denoted by z_t. It is also assumed that $[e_t, u_t, x_t, z_t]'$, $t=1,2,\ldots,T$, are distributed as independent drawings from a multivariate normal distribution with mean $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Sigma}$, where

$$\underset{\sim}{\Sigma} = \begin{pmatrix} \sigma_e^2 & \sigma_{ue} & 0 & 0 \\ \sigma_{ue} & \sigma_u^2 & 0 & 0 \\ 0 & 0 & \sigma_x^2 & \sigma_{xz} \\ 0 & 0 & \sigma_{xz} & \sigma_z^2 \end{pmatrix}, \qquad (2.3)$$

$\sigma_{xz} \neq 0$ and $\sigma_x^2 > 0$. We note that the model as defined is equivalent to the system of structural equations

$$\begin{aligned} Y_t &= \beta X_t + w_t \\ X_t &= \delta z_t + \epsilon_t , \end{aligned} \qquad (2.4)$$

where Y_t and X_t are endogenous variables, z_t is an exogenous variable, and $\delta = \sigma_{xz}/\sigma_z^2$.

The model is easily generalized to contain an intercept term and the assumption of zero means for all variables is an assumption of convenience. The assumption of normality is stronger than necessary for many of the results used in this paper, but simplifies the presentation and is the model used in our Monte Carlo study and that of Feldstein [4].

Under certain circumstances it may be known that $\sigma_{ue}=0$. This situation is identified as model b). The situation wherein σ_{ue} is unknown and unrestricted is model a).

3. ESTIMATION

3.1. Estimation when σ_{ue} is unknown (model a)

The reduced form associated with (2.4) is given by

$$\begin{aligned} Y_t &= \beta\delta z_t + \xi_t , \\ X_t &= \delta z_t + \epsilon_t , \end{aligned} \qquad (3.1)$$

where ξ_t and ϵ_t are the reduced form errors. We assume $\sigma_\epsilon^2 > 0$ for model a). For this model the maximum likelihood estimator of β is the instrumental variable estimator,[1]

$$\tilde{\beta}_{IV} = \left(\sum_{t=1}^{T} X_t z_t\right)^{-1} \sum_{t=1}^{T} Y_t z_t . \qquad (3.2)$$

Under the model assumptions, it is well known that

$$T^{1/2}(\tilde{\beta}_{IV}-\beta) \xrightarrow{\mathcal{L}} N(0,[\delta^2\sigma_z^2]^{-1}\sigma_w^2) . \qquad (3.3)$$

See, for example, Fuller [5].

The matrix of sums of squares and cross products of residuals from the reduced form regressions is

$$\underset{\sim}{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix} ,$$

where

$$W_{11}=\Sigma\hat{\xi}^2=\Sigma Y^2-[\Sigma z^2]^{-1}(\Sigma Yz)^2 , \qquad (3.5)$$

$$W_{12}=\Sigma\hat{\xi}\hat{\epsilon}=\Sigma XY-[\Sigma z^2]^{-1}(\Sigma Xz)(\Sigma Yz) \qquad (3.6)$$

$$W_{22}=\Sigma\hat{\epsilon}^2=\Sigma X^2-[\Sigma z^2]^{-1}(\Sigma Xz)^2 . \qquad (3.7)$$

The family of modified limited information maximum likelihood estimators of β considered by Fuller [5] is defined by

$$\tilde{\beta}_{ML\alpha} = \frac{(T-1)(\Sigma Yz)(\Sigma Xz) + \alpha W_{12}(\Sigma z^2)}{(T-1)(\Sigma Xz)^2 + \alpha W_{22}(\Sigma z^2)} , \qquad (3.8)$$

where α is a fixed positive real number.

The ordinary least squares estimator of β is $\hat{\beta}_{OLS} = (\Sigma X^2)^{-1}\Sigma XY$. When using the mean square error as the basis of comparison, neither the ordinary least squares nor the instrumental variable estimator is preferred for all parameter values.

We now construct an estimator for β of the form (1.1) with weights (1.2). In our construction $\tilde{\beta}_{ML4}$ plays the role of $\hat{\theta}_1$ and $\tilde{\beta}_R=W_{22}^{-1}W_{12}$, where W_{22} and W_{12} are defined in (3.7) and (3.6), plays the role of $\hat{\theta}_2$. Our randomly weighted average estimator is

$$\tilde{\beta}_W = \tilde{a}_W\tilde{\beta}_{ML4} + (1-\tilde{a}_W)\tilde{\beta}_R, \qquad (3.15)$$

where

$$\tilde{a}_W = \frac{(\tilde{\beta}_{ML4} - \tilde{\beta}_R)^2 + \hat{A}_2}{\hat{A}_1 + \hat{A}_2 + (\tilde{\beta}_{ML4} - \tilde{\beta}_R)^2} ,$$

$$\hat{A}_2 = \frac{(T-1)S_1^2}{(T-5)W_{22}} ,$$

$$S_1^2 = (T-2)^{-1}(W_{11} - W_{12}\tilde{\beta}_R) .$$

The following theorem demonstrates that $T^{1/2}(\tilde{\beta}_W-\beta)$ converges in distribution to the estimator studied by Huntsberger.

Theorem 3.1.1. Let $H(\theta,\gamma,A_1,A_2)$ denote the distribution of the randomly weighted average estimator with weight (1.2).

Let (X,Y,z) be distributed as a trivariate normal random variable satisfying model (2.1), (2.2), (2.3), (2.4) with $\delta \neq 0$ and $\sigma_x^2 - \delta^2\sigma_z^2 > 0$. Let all parameters except σ_u^2 and σ_{ue} be fixed. Let $\sigma_u^2 = \alpha T^{-1/2}$ and $\sigma_{ue} = \eta\sigma_u^2$ where α and η are fixed. Let $\tilde{\beta}_W$ be defined by (3.15). Then

$$T^{1/2}(\tilde{\beta}_W-\beta) \xrightarrow{\mathcal{L}} H(0,\gamma,A_1,A_2) ,$$

where

$$\gamma^2 = \frac{\alpha^2(\eta-\beta)^2\delta^2\sigma_z^2}{\sigma_e^2\sigma_x^2(\sigma_x^2-\delta^2\sigma_z^2)}$$

$$A_1 = (\delta^2\sigma_z^2)^{-1}\sigma_e^2$$

---

[1] All summations in this paper are over t as t ranges from 1 to T. Henceforth, we shall suppress the subscripts and the range of summation.

$$A_2 = (\sigma_x^2 - \delta^2\sigma_z^2)^{-1}\sigma_e^2 .$$

3.2. Estimation when σ_{ue} is known (model b)

By assumption, the instrumental variable z is correlated with x and uncorrelated with u and e. Thus, we may write

$$z_t = \rho x_t + v_t , \tag{3.18}$$

where v_t, t = 1,2,...,T are normal independent $(0,\sigma_v^2)$ random variables independent of u_j, e_j and x_j for all t and j. Given our model, with the added assumption that $\sigma_{ue} = 0$, the vector (Y_t, X_t, z_t) is normally distributed with zero mean and covariance matrix $\underset{\sim}{V}$, where

$$\underset{\sim}{V} = \begin{pmatrix} \beta^2\sigma_x^2+\sigma_e^2 & \beta\sigma_x^2 & \rho\beta\sigma_x^2 \\ \beta\sigma_x^2 & \sigma_x^2+\sigma_u^2 & \rho\sigma_x^2 \\ \rho\beta\sigma_x^2 & \rho\sigma_x^2 & \rho^2\sigma_x^2+\sigma_v^2 \end{pmatrix} . \tag{3.19}$$

We shall obtain the maximum likelihood estimator of $\underset{\sim}{\theta} = (\beta, \sigma_x^2, \rho, \sigma_e^2, \sigma_u^2, \sigma_v^2)'$ under the assumptions; $\beta \neq 0$, $\rho \neq 0$, $\sigma_x^2 \neq 0$, $\sigma_e^2 \geq 0$, $\sigma_u^2 \geq 0$, $\sigma_v^2 \geq 0$. The inequality restrictions on β, ρ and σ_x^2 are required for all parameters of the model to be identified. Once the maximum likelihood estimator of $\underset{\sim}{\theta}$ has been obtained we shall demonstrate that, given the remaining assumptions, the estimator of β is consistent for all β, including $\beta = 0$. We also temporarily assume that at most one of the variances σ_e^2, σ_u^2 or σ_v^2 is zero. If two or more of the population error variances are zero, then the matrix $\underset{\sim}{V}$, defined in (3.19), is singular and the vector (Y,X,z)' has a singular normal distribution. This situation is easily detected in the sample because the matrix of sums of squares and cross-products of (Y,X,z)' is singular. Therefore, this case will be treated separately. Under the present assumptions, the space of admissible values for the parameter vector, $\underset{\sim}{\theta}$, is denoted by Θ.

If there are no restrictions on the matrix $\underset{\sim}{V}$, the maximum likelihood estimators of the elements of $\underset{\sim}{V}$ are given by the sample moments, $\hat{V}_{11} = s_X^2 = T^{-1}\Sigma X^2$, $\hat{V}_{22} = s_Y^2 = T^{-1}\Sigma Y^2$, $\hat{V}_{12} = s_{XY} = T^{-1}\Sigma XY$, $\hat{V}_{13} = s_{Yz} = T^{-1}\Sigma Yz$ and $\hat{V}_{23} = s_{Xz} = T^{-1}\Sigma Xz$ (see [1], Chap. 3). The simple estimator of the parameter vector, $\underset{\sim}{\theta}$, is obtained by solving the equations $\underset{\sim}{V} = \underset{\sim}{\hat{V}}$ for the parameters of interest. The following theorem defines the maximum likelihood estimator of $\underset{\sim}{\theta}$ (i.e. the estimator that maximizes the likelihood on the parameter space Θ).

Theorem 3.2.1. Let (Y,X,z) be distributed as a trivariate normal random variable with mean zero and covariance matrix $\underset{\sim}{V}$, defined in (3.19). Let the parameter $\underset{\sim}{\theta} \in \Theta$ and let G denote the event that the simple estimator does not lie in Θ. Let $\hat{R}_{XY}^2 = (s_Y^2 s_X^2)^{-1} s_{XY}^2$, $\hat{R}_{Xz}^2 = (s_X^2 s_z^2)^{-1} s_{Xz}^2$, $\hat{R}_{Yz}^2 = (s_Y^2 s_z^2)^{-1} s_{Yz}^2$ and $R_{min}^2 = \min\{\hat{R}_{XY}^2, \hat{R}_{Xz}^2, \hat{R}_{Yz}^2\}$ and define

$$\begin{aligned} \hat{\underset{\sim}{\theta}}_e &= (\hat{\beta}_{IN}, s_Y^{-2}s_{XY}^2, s_{XY}^{-1}s_{Yz}, 0, s_X^2 - s_Y^{-2}s_{XY}^2, s_z^2 - s_Y^{-2}s_{Yz}^2)' , \\ \hat{\underset{\sim}{\theta}}_u &= (\hat{\beta}_{OLS}, s_X^2, s_X^{-2}s_{Xz}, s_Y^2 - s_X^{-2}s_{XY}^2, 0, s_z^2 - s_X^{-2}s_{Xz}^2)' , \\ \hat{\underset{\sim}{\theta}}_v &= (\tilde{\beta}_{IV}, s_z^{-2}s_{Xz}^2, s_{Xz}^{-1}s_z^2, s_Y^2 - s_z^{-2}s_{Yz}^2, s_X^2 - s_z^{-2}s_{Xz}^2, 0)' , \\ \hat{\beta}_{IN} &= s_{XY}^{-1}s_Y^2 . \end{aligned} \tag{3.29}$$

Then the maximum likelihood estimator of $\underset{\sim}{\theta}$ is

$$\hat{\theta}_M = \begin{cases} \hat{\underset{\sim}{\theta}}_e, & \text{if G and } R_{min}^2 = \hat{R}_{Xz}^2 \\ \hat{\underset{\sim}{\theta}}_u, & \text{if G and } R_{min}^2 = \hat{R}_{Yz}^2 \\ \hat{\underset{\sim}{\theta}}_v, & \text{if G and } R_{min}^2 = \hat{R}_{XY}^2 \\ \hat{\underset{\sim}{\theta}}, & \text{otherwise.} \end{cases} \tag{3.30}$$

That is, $\hat{\underset{\sim}{\theta}}_M \in \Theta$ and $L(\hat{\theta}_M) \geq L(\underset{\sim}{\theta}^+)$ for all $\underset{\sim}{\theta}^+ \in \Theta$.

If exactly two of the variances are zero, then there is a perfect correlation between the two variables with no measurement error. In such a case the model can be reduced to a regression model in one of these variables and the third variable. The maximum likelihood estimator for the slope parameter of the reduced model is obtained by ordinary least squares. The maximum likelihood estimator was constructed under the assumption that the parameter β is not zero. It can be proven that the first component of $\hat{\underset{\sim}{\theta}}_{ML}$ is consistent for β when $\beta = 0$.

Recall that the modified estimator, $\tilde{\beta}_{ML4}$, has moment properties superior to those of $\tilde{\beta}_{IV}$. The estimator, $\hat{\beta}_{IN}$, defined in (3.29) can be modified in the same way to produce the estimator,

$$\hat{\beta}_{IN4} = \frac{(T-1)\Sigma XY \Sigma Y^2}{(T-5)(\Sigma XY)^2 + 4\Sigma Y^2 \Sigma X^2} ,$$

with moment properties superior to $\hat{\beta}_{IN}$. On the basis of these results and Theorem 3.2.1, we propose the following estimator for β when σ_{ue} is known to be zero:

$$\tilde{\beta}_{MA} = \begin{cases} \hat{\beta}_{OLS}, & \text{if G and } R_{min}^2 = \hat{R}_{Yz}^2, \text{ or } \sigma_u^2 = \sigma_e^2 = 0, \\ & \text{or } \sigma_u^2 = \sigma_v^2 = 0 \\ \hat{\beta}_{IN4}, & \text{if G and } R_{min}^2 = \hat{R}_{Xz}^2, \text{ or } \sigma_e^2 = \sigma_v^2 = 0, \\ & \sigma_u^2 \neq 0 \\ \tilde{\beta}_{ML4}, & \text{otherwise.} \end{cases} \tag{3.36}$$

A weighted average estimator analogous to (3.36) is

$$\tilde{\beta}_{WA} = \begin{cases} \hat{\beta}_{OLS}, & \text{if G and } R_{min}^2 = \hat{R}_{Yz}^2, \text{ or } \sigma_u^2 = \sigma_e^2 = 0, \\ & \text{or } \sigma_u^2 = \sigma_v^2 = 0 \\ \hat{\beta}_{IN4}, & \text{if G and } R_{min}^2 = \hat{R}_{Xz}^2, \text{ or } \sigma_e^2 = \sigma_v^2 = 0, \\ & \sigma_u^2 \neq 0 \\ \tilde{\beta}_W , & \text{otherwise,} \end{cases} \tag{3.39}$$

where $\tilde{\beta}_W$ is defined in (3.15).

An estimator patterned after the arguments of Feldstein[1] is

$$\tilde{\beta}_K = \hat{\eta}\,\hat{\beta}_{OLS} + (1-\hat{\eta})\tilde{\beta}_{IV} \qquad (3.40)$$

where

$$\hat{\eta} = [1+(T+1)\hat{K}^2\{H_1(\hat{K}^2+H_2)+2\hat{K}^2\}^{-1}]^{-1}$$

$$H_1 = s_{Xz}^{-2}(s_X^2 s_z^2 - s_{Xz}^2)$$

$$H_2 = s_{XY}^{-2}(s_X^2 s_Y^2 - s_{XY}^2)$$

$$\hat{K} = \begin{cases} (\hat{\beta}_{OLS})^{-1}\tilde{\beta}_{IV}-1 & \text{if } (\hat{\beta}_{OLS})^{-1}\tilde{\beta}_{IV} > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Feldstein defined $\hat{K}$ to be an estimator of $K = \sigma_x^{-2}\sigma_u^2$. For details the reader is referred to [4].

4. EMPIRICAL RESULTS

Simulation experiments were carried out to compare the estimators $\hat{\beta}_{OLS}$, $\tilde{\beta}_{ML1}$, $\tilde{\beta}_{ML4}$, $\tilde{\beta}_W$, $\tilde{\beta}_{MA}$, $\tilde{\beta}_{WA}$ and $\tilde{\beta}_K$. Our study follows closely that of Feldstein [4]. As in that study we considered two cases, T = 25 and T = 100, with $\beta = 1$, $\sigma_{ue} = 0$ and $\sigma_x^2 = \sigma_z^2 = 100$. The remaining parameters, σ_u^2, σ_e^2 and σ_{xz}, were also chosen to coincide with Feldstein's study. Two thousand samples were generated and the seven estimators were computed. The random number package, "Super Duper," from McGill University was used to generate values for the normal random variables e_t, u_t, x_t and z_t.

Three conclusions are possible from the Monte Carlo results for estimators that do not use the information $\sigma_{ue} = 0$. These are:

1) The modified limited information estimator with $\alpha=4$ has mean square error which is smaller than that with $\alpha=1$ for all parameter sets.
2) The modified limited information estimator with $\alpha=1$ is very nearly unbiased for all values of the parameters.
3) The estimator $\tilde{\beta}_W$ behaves like the randomly weighted average estimator discussed by Huntsberger. Loosely speaking, the mean square error of $\tilde{\beta}_{ML4}$ is larger than that of $\tilde{\beta}_W$ when $\gamma < 1.17$. As γ increases the relative mean square error of $\tilde{\beta}_W$ reaches a maximum of about 1.2 for γ between 2.0 and 3.0. As γ becomes larger the mean square error of $\tilde{\beta}_W$ approaches that of $\tilde{\beta}_{ML4}$.

The superiority of $\tilde{\beta}_{ML4}$ over $\tilde{\beta}_{ML1}$ was greater for the smaller correlation between X and z. This was expected because the two estimators are identical if $R_{Xz} = 1$.

Similar conclusions are reached when the estimators using the information that $\sigma_{ue}=0$ are compared to that of $\tilde{\beta}_{MA}$. The ratio of the mean square error of $\tilde{\beta}_{WA}$ is less than one for $\gamma=0$, increases to a peak above one and then approaches one from above as γ increases. Therefore, the comments made about the pair $\tilde{\beta}_{ML4}$ and $\tilde{\beta}_W$ also apply to the pair $\tilde{\beta}_{MLA}$ and $\tilde{\beta}_{WA}$. The mean square error of $\tilde{\beta}_K$ is uniformly larger than that of $\tilde{\beta}_{WA}$. The superiority of $\tilde{\beta}_{WA}$ can be attributed to two factors. First, $\tilde{\beta}_{ML4}$ performs uniformly better than the instrumental variable estimator, $\tilde{\beta}_{IV}$. Thus, a randomly weighted average of $\tilde{\beta}_{ML4}$ and $\tilde{\beta}_{OLS}$ is expected to perform better than a randomly weighted average of $\tilde{\beta}_{IV}$ and $\hat{\beta}_{OLS}$. Second, $\tilde{\beta}_{WA}$ is restricted so that the estimators for σ_x^2, σ_u^2, σ_e^2 and σ_v^2 are nonnegative whereas $\tilde{\beta}_K$ only guarantees a nonnegative estimator for $\sigma_x^{-2}\sigma_u^2$.

When the estimators using the information $\sigma_{ue}=0$ are compared to those that do not, the estimators using the information are generally superior.

5. CONCLUSION

On the basis of this study Fuller's modified limited information estimator with $\alpha=4$ can be recommended for instrumental variable estimation when σ_{ue} is unknown and the objective is to minimize the mean square error of the estimator of β. Likewise, the adjusted maximum likelihood estimator, $\tilde{\beta}_{MA}$, can be recommended if σ_{ue} is known to be zero. The mean square error function of the randomly weighted estimators is similar to that obtained by Huntsberger for analogous combinations of normal estimators. That is, the randomly weighted estimator has a smaller mean square error than the modified maximum likelihood estimator if the bias in the second estimator is 'small' and a larger mean square error otherwise.

Proof of the theorems and tables of the Monte Carlo results are contained in a larger manuscript which can be obtained from the authors.

ACKNOWLEDGMENT

This research was partly supported by the Bureau of the Census through Joint Statistical Agreements J.S.A. 75-1, and J.S.A. 76-66.

[1]The estimator is not identical to that considered by Feldstein.

REFERENCES

[1] Anderson, T.W., An Introduction to Multivariate Statistical Analysis, New York: John Wiley, 1958.

[2] Cochran, W.G., "Errors of Measurement in Statistics," Technometrics, 10 (November 1968), 637-666.

[3] Durbin, J., "Errors in Variables," Revue de L'Institut International de Statistique, 22, No. 1/3 (1954), 23-32.

[4] Feldstein, M.S., "Errors in Variables: A Consistent Estimator with Smaller MSE in Finite Samples," Journal of the American Statistical Association, 69 (December 1974), 990-996.

[5] Fuller, W.A., "Some Properties of a Modification of the Limited Information Estimator," to appear in Econometrica.

[6] Fuller, W.A., "Properties of Some Estimators for the Errors-in-Variables Model," presented at the New Orleans meeting of the Econometric Society (December 1971).

[7] Geary, R.C., "Inherent Relations Between Random Variables," Proceedings of the Royal Irish Academy, Vol. 47, Sect. A, No. 6 (1942).

[8] Goldberger, A.S., Econometric Theory, New York: John Wiley, 1964.

[9] Hansen, M.H., Hurwitz, W.N. and Madow, W.G., Sample Survey Methods and Theory, Vol. 2, New York: John Wiley, 1953.

[10] Huntsberger, D.V., "A Generalization of a Preliminary Testing Procedure for Pooling Data," Annals of Mathematical Statistics, 26 (December 1955), 734-743.

[11] Johnston, J., Econometric Methods, New York: McGraw-Hill, 1963.

[12] Kendall, M.G. and Stuart, A., The Advanced Theory of Statistics Vol. 2, New York: Hafner Publishing Company, 1967.

[13] Malinvaud, E., Statistical Methods of Econometrics, Amsterdam: North-Holland Publishing Company, 1966.

[14] Reiersol, O., "Confluence Analysis by Means of Instrumental Sets of Variables," Arkiv fur Mathematik, Astronomi och Fysik, 32A, 4 (1945).

[15] Sargan, J.D., "The Estimation of Economic Relationships Using Instrumental Variables," Econometrica, 26 (July 1958), 393-415.

[16] Sargan, J.D. and Mikhail, W.M., "A General Approximation to the Distribution of Instrumental Variables Estimates," Econometrica, 39 (January 1971), 131-169.

ON ESTIMATION OF POPULATION PROPORTIONS BY DIRECT AND INVERSE SAMPLING

R. P. Chakrabarty and P. J. Tsai
Jackson State University and University of Georgia

Summary

In this paper, we compare the relative merits of four sampling methods, (1) Direct sampling without replacement, (2) Direct sampling with replacement, (3) Inverse sampling without replacement and (4) Inverse sampling with replacement that may be used to estimate population proportions. The stability of the variance estimator in addition to the efficiency of the estimator is taken into consideration in judging the performance of an estimator. Conditions under which one method is preferable to another are obtained. Sample size needed for the use of asymptotic results are indicated by simulation.

1. INTRODUCTION

In socio-economic and ecological surveys, one is frequently interested in estimating either the number NP or proportion P which possesses a specified characteristic in a finite population of size N. The unbiased estimator of the proportions, its variance and the data-base variance estimator for direct sampling with and without replacement are well known (see Cochran 1963). Haldane (1945) first obtained an unbiased estimator of P and its asymptotic variance for the case of inverse sampling with replacement. Recently Best (1974) obtained a closed form for the variance of Haldane's unbiased estimator of the P and compared its efficiency with that of the maximum likelihood estimator given by Feller (1968). Best's numerical evaluation of efficiency showed that the unbiased estimator of P was generally more efficient than the maximum likelihood estimator. Scheaffer (1974) compared the unbiased estimator in direct sampling with replacement and the biased estimator in inverse sampling with replacement, and obtained the asymptotic stabilities of the variance estimators. He has shown that under certain conditions inverse sampling with replacement provides a more stable variance estimator in large samples.

It may be noted that the results of Haldane, Feller, Best and Scheaffer were obtained assuming infinite populations. Some new results for the important case of sampling without replacement from a finite population are obtained in this paper. Only unbiased estimators are considered in view of Bests' results. Special emphasis has been given to obtain results that are exact for any sample size whenever it is possible. We note that the estimate of the variance of an estimator is often used in drawing statistical inferences (e.g., confidence limits for the parameters). It is, therefore, desirable that the variance estimator should be as stable as possible. The importance of investigation of the stability of the variance estimator was pointed out by Chakrabarty and Rao (1967) and Chakrabarty (1973) in estimation of ratios. We, therefore, consider the stability of the variance estimator associated with an estimator in addition to the efficiency of the estimator itself in judging the performance of an estimator. Without loss of generality, we shall discuss the problem of estimation of the proportion P only in a finite population of N units.

Some notations are introduced here that are used in subsequent sections:

- N : number of units in the population.
- N_1 : number of units in the population that possesses a specified attribute or characteristic or falls into specified class.
- n : total sample size.
- n_1 : sample size from the specified class.
- P : $P = N_1 / N$ population proportion to be estimated.
- Q : $Q = 1 - P$.
- p_1 : the unbiased estimator of P by direct sampling with replacement; DWR.
- p_2 : the unbiased estimator of P by direct sampling without replacement; DWTR.
- p_3 : the unbiased estimator of P by inverse sampling with replacement; IWR.
- p_4 : the unbiased estimator of P by inverse sampling without replacement; IWTR.
- q_i : $q_i = 1 - p_i$, i = 1, 2, 3, 4.
- f : sampling fraction, $f = \frac{n}{N}$.
- c : finite population correction factor.
- c : $1 - f = 1 - n/N$.
- $V(p_i)$: the variance of p_i, i = 1, 2, 3, 4.
- $v(p_i)$: the estimator of $V(p_i)$, i = 1, 2, 3, 4.
- $V_e(v(p_i))$: the exact variance of $v(p_i)$, i = 1, 2.
- $V_a(v(p_i))$: asymptotic variance of $v(p_i)$, i = 1, 2, 3, 4.

2. DIRECT SAMPLING

2.1 Direct Sampling with Replacement; DWR. The probability density function (p.d.f.) of n_1 is

$$f(n_1) = \binom{n}{n_1} P^{n_1} Q^{n-n_1} \quad ; \quad n_1 = 0, 1, 2, \ldots n$$

$$P + Q = 1 \quad , \quad P, Q \geqslant 0$$

It is well known that the unbiased estimator of P, its variance and unbiased variance estimator are given by

$$p_1 = n_1/n$$

$$V(p_1) = PQ/n$$

and

$$v(p_1) = p_1 (1-p_1) / (n-1) \qquad (2.1)$$

respectively (see Cochran).

It can be shown that the exact variance of $v(p_1)$ is given by

$$V_e(v(p_1)) = (A_1 + A_2 + A_3 + A_4) / (n^4(n-1)^2)$$

where

$$A_1 = m_1^{(4)} - 2(n-3)m_1^{(3)}$$

$$A_2 = (n^2 - 6n + 7)m_1^{(2)}$$

$$A_3 = (n-1)^2 m_1^{(1)}$$

$$-A_4 = \left\{(n-1)m_1^{(1)} - m_1^{(2)}\right\}^2 .$$

$$m_1^{(k)} = E(n_1^{(k)}) = E\left[n_1(n_1-1)\ldots(n_1-k+1)\right];$$

$$k = 1, 2, 3, 4 \qquad (2.2)$$

The asymptotic variance of $v(p_1)$ by using the well known formula (Kendall and Sturat, 1968)

$$V(g(x)) \doteqdot g'(E(x)) \; V(x) \quad ,$$

is obtained as

$$V_a(v(p_1)) = \left\{v'(p)\right\}^2 V(p_1) \qquad (2.3)$$

$$\simeq (1-2P)^2 PQ/(n(n-1)^2)$$

2.2 Direct Sampling without Replacement; DWTR

The p.d.f. of n_1 is

$$f(n_1) = \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}},$$

$$n_1 = 0, 1, \ldots\ldots, \min(n, N_1) \qquad (2.4)$$

The unbiased estimator of P, its variance and unbiased variance estimator are given by:

$$p_2 = n_1/n$$

$$V(p_2) = PQ(1-(n-1)/N-1))/n$$

and $v(p_2) = p_2 q_2 (1-n/N)/(n-1) \qquad (2.5)$

respectively. Using the relations between moments about origin and descending factorial moments of n_1 for the p.d.f. (2.4) it can be shown that the exact variance of $v(p_2)$ is given by:

$$V_e(v(p_2) = C^2 (B_1+B_2+B_3+B_4)/(n^4 (n-1)^2)$$

where

$$B_1 = m_2^{(4)} - 2(n-3)m_2^{(3)}$$

$$B_2 = (n^2 - 6n + 7) m_2^{(2)}$$

$$B_3 = (n-1)^2 m_2^{(1)}$$

$$B_4 = -((n-1)m_2^{(1)} - m_2^{(2)}).$$

$$m_2^{(k)} = E(n_1^{(k)}) = n^{(k)} N_1^{(k)} / N^{(k)};$$

$$k = 1, 2, 3, 4 \qquad (2.6)$$

The asymptotic variance of $v(p_2)$ is given by

$$V_a(v(p_2) = (\frac{c}{(n-1)} (1-2P))^2 V(p_2)$$

$$= (1-2P)^2 PQ (1-n/N)^2 (1-(n-1)/$$

$$(N-1)) / (n(n-1)^2) \qquad (2.7)$$

3. INVERSE SAMPLING

Inverse Sampling is a procedure where sampling is continued until n_1 units possessing the specified characteristic have been obtained. The number n_1 is pre-determined and the Sample Size $n(\geqslant n_1)$ is thus a random variable.

3.1. Inverse Sampling with Replacement: IWR. The p.d.f of n is

$$f(n) = \binom{n-1}{n_1-1} P^{n_1} Q^{n-n_1} \quad ; \; n=n_1, n_1+1, \ldots \qquad (3.1)$$

Haldane (1945) first gave the unbiased estimator of P and its variance as

$$P_3 = (n_1-1)/(n-1)$$

and $V(p_3) = \frac{P^2Q}{n_1}\left(1 + \frac{2!Q}{n_1+1} + \frac{3!Q^2}{(n_1+1)(n_1+2)} + \ldots\right)$ (3.2)

respectively. Note that

$E(n) = n_1 / P$ and $V(n) = n_1 Q/P^2$ (3.3)

consequeltly, the asymptotic variance of p_3 is

$$V(p_3) \doteq V(n)\left[(n_1-1)^2 / (E(n)-1)^4\right]$$
$$= n_1 (n_1-1)^2 P^2 Q / (n_1 - P)^4 \quad (3.4)$$

Now, $E\left[\frac{(n_1-1)(n_1-2)}{(n-1)(n-2)}\right] = \sum_{n=n_1}^{\infty} \frac{(n_1-1)(n_1-2)}{(n-1)(n-2)} \times \binom{n-1}{n_1-1} P^{n_1} Q^{n-n_1} = P^2$

$$E\left[\left(\frac{n_1-1}{n-1}\right)^2 - \frac{(n_1-1)(n_1-2)}{(n-1)(n-2)}\right]$$
$$= E(p_3^2) - E^2(p_3) = V(p_3)$$

Thus
$$v(p_3) = \left(\frac{n_1-1}{n-1}\right)^2 - \frac{(n_1-1)(n_1-2)}{(n-1)(n-2)}$$
$$= \frac{(n_1-1)(n-n_1)}{(n-1)^2(n-2)} = \frac{p_3(1-p_3)}{(n-2)} \quad (3.5)$$

is an unbiased estimator of $V(p_3)$. The asymptotic variance of $v(p_3)$ is obtained by using (2.3) as

$$Va(v(p_3)) = \left[n_1 (n_1-1)^2\left\{(3P-2)n_1^2 + (2-5P)n_1 P + 2P^2\right\}^2 P^4 Q\right] / \left[(n_1-P)^6 (n_1-2P)^4\right] \quad (3.6)$$

3.2 Inverse Sampling without Replacement; IWTR

The p.d.f. of sample size n is

$$f(n) = \frac{\binom{N_1}{n_1-1}\binom{N-N_1}{n-n_1}}{\binom{N}{n-1}} \cdot \frac{N_1-n_1+1}{N-n+1};$$

$n = n_1, n_1+1, \ldots\ldots. N$ (3.7)

The kth ascending factorial moment of n is $m_4^{[k]} = E(n^{[k]}) = n_1^{[k]} (N+1)^{[k]} / (N_1+1)^{[k]}$

Where $n^{[k]} = n(n+1)(n+2)\ldots\ldots(n+k-1)$ (3.8)

consequently, $E(n) = n_1 (N+1) / (N_1+1)$ (3.9)

and $V(n) = \frac{n_1(N+1)(N-N_1)(N_1-n_1+1)}{(N_1+1)^2(N_1+2)}$

or

$$V(n) = \frac{n_1\left(1+\frac{1}{N}\right)\left(P - \frac{n_1-1}{N}\right)Q}{\left(P+\frac{1}{N}\right)^2\left(P+\frac{2}{N}\right)} \quad (3.10)$$

Since

$$E\left(\frac{n_1-1}{n-1}\right) = \sum_{n=1}^{N} \frac{(n_1-1)\binom{N_1}{n_1-1}\binom{N-N_1}{n-n_1}}{(n-1)\binom{N}{n-1}}.$$

$$\frac{N_1-n_1+1}{N-n+1} = \frac{N_1}{N} = P$$

$$p_4 = \frac{n_1-1}{n-1} \quad (3.11)$$

is an unbiased estimator of P.

The asymptotic variance of p_4 is

$$V(p_4) \doteq \frac{(n_1-1)^2}{\left\{E(n)-1\right\}^4} V(n) \quad (3.12)$$

where E(n) and V(n) are given in (3.9) and (3.10) respectively.

We show that an unbiased estimator of the exact variance of p_4 can be obtained even though an explicit expression for $V(p_4)$ could not be obtained. By definition, we have

$$V(p_4) = E(p_4^2) - P^2.$$

Since

$$\left(\frac{n_1-1}{n-1}\right)^2 - \frac{(n_1-1)(n_1-2)}{(n-1)(n-2)} = \frac{p_4 q_4}{n-2}$$

and $E((n_1-1)(n_1-2)/(n-1)(n-2)) = P(N_1-1)/(N-1)$

We get $E(p_4q_4/(n-2)) = V(p_4) + PQ/(N-1)$

Further, $E(p_4q_4) = PQ - V(p_4)$

Consequently, $v(p_4) = (p_4q_4)(1 - (n-1)/N)/(n-2)$

$$\tag{3.13}$$

is an unbiased estimator of $V(p_4)$.

Now, the derivative of $v(p_4)$ with respect to n is

$$\frac{dv(p_4)}{dn} = (n_1-1)\left(\frac{-2n^2 + (2+3n_1)n + 2 - 5n_1}{(n-1)^3(n-2)^2}\cdot\left(1 - \frac{n-1}{N}\right) - \frac{n-n_1}{(n-1)^2(n-2)}\cdot\frac{1}{N}\right)$$

$$= g(n) \quad \text{(say)} \tag{3.14}$$

Accordingly, the asymptotic variance of $v(p_4)$ is

$$V_a(v(p_4)) \doteq (g(E(n)))^2 V(n)$$

where E(n) and V(n) are given in (3.9) and (3.10) respectively. The above expression is too complicated for analytical comparison but can be easily evaluated numerically.

4. COMPARISON OF DIRECT AND INVERSE SAMPLING

Any comparison of direct sampling with a fixed sample size n to inverse sampling with variable sample size n would be made on the basis of expected sample size, E(n). Note that in

IWR: $E(n) = n_1/P$ or $n_1 = P.E(n)$

IWTR: $E(n) = n_1(1+1/N)/(P+1/N)$

$$= n_1/P \quad \text{for large N} \tag{4.1}$$

Therefore, n_1 is replaced by P.E(n) in variance formulas in inverse sampling for comparison with variance formulas in direct sampling. The results are given in the sequal.

4.1. Efficiencies of Unbiased Estimators pi

It can be shown that for large n or E(n)

$$V(p_1) = PQ/n \quad , \quad V(p_2) \doteq PQ(1-n/N)/n$$

$$V(p_3) \doteq PQ/E(n) \quad , \quad V(p_4) \doteq PQ(1-E(n)/N)/E(n) \tag{4.2}$$

Thus p_2 in DWTR is more efficient than p_1 in DWR and p_4 in IWTR is more efficienct than p_3 in IWR. But p_1 and p_3 as well as p_2 and p_4 are equally efficient in large samples.

Sampling without replacement is definitely preferable but one cannot choose between DWTR and IWTR on the basis of the efficiency of unbiased estimators of P.

4.2. Stabilities of Variance Estimators v(pi)

Following Rao and Chakrabarty (1967) and Chakrabarty (1973) the stability of the variance estimator $v(p_i)$ relative to $v(p_j)$ is given by

$$(CV(v(p_i)))^2/(CV(v(p_j)))^2 \tag{4.3}$$

where $(CV(v(p_i)))^2 = V(v(p_i))/(V(p_i))^2$ i = 1,2,3,4 is the relative variance of $v(p_i)$ since $E(v(pi)) = V(p_i)$. The variance estimator $v(p_i)$ is more stable than $v(p_j)$ if $(CV(v(p_i)))^2 < (CV(v(p_j)))^2$

Now from (2.2), (2.4), (2.6), and (2.7) it can be shown that $Ve(v(p_1)) > Ve(v(p_2))$ and $Va(v(p_1)) > Va(v(p_2))$ for all P.

Substituting n_1 by P.E(n) in variance formulas in inverse sampling and equating E(n) in inverse sampling to n in direct sampling we obtain after some algebraic manipulations the conditions under which one variance estimator is more stable than another. The details which are given in a technical report by the authors (1976) are omitted here and only the results are summarized below:

(1) $v(p_2)$ in DWTR is more stable than $v(p_1)$ in DWR for all P.

(2) $v(p_3)$ in IWR is more stable than $v(p_1)$ in DWR iff $P > 0.6$.

(3) $v(p_4)$ in IWTR is more stable than $v(p_1)$ in DWR iff

$$P > \frac{2-d-(1-d)^{\frac{1}{2}}}{3-2d+2(1-d)^{\frac{1}{2}}}$$

(4) $v(p_3)$ in IWR is more stable than $v(p_2)$ in in DWTR iff

$$\frac{2-(1-d)^{\frac{1}{2}}}{3+2(1-d)^{\frac{1}{2}}} < P < \frac{2+(1-d)^{\frac{1}{2}}}{3-2(1-d)^{\frac{1}{2}}}$$

(5) $v(p_4)$ in IWTR is more stable than $v(p_2)$ in DWTR iff

$$P > \frac{3-2d}{5-4d}$$

(6) $v(p_4)$ in IWTR is more stable than $v(p_3)$ in IWR iff

$$P > \frac{2-d+2(1-d)^{\frac{1}{2}}}{3-2d+3(1-d)^{\frac{1}{2}}}.$$

where $d=n/N$

These results are shown graphically indicating the regions of stable variance estimator. The implication of these results are that if we have prior knowledge about the range of the proportion parameter P, then it is possible to choose the optimum sampling techniques from the four techniques we have discussed, particularly between DWTR and IWTR.

We may mention that we have also evaluated numerically the stabilities of the variance estimators by generating 1000 samples, each of size 100 from a population with N=1000 and P=.8 on cyber 74. DWTR and IWTR was equally efficient but IWTR provided the most stable variance estimator.

We have also investigated the sample size needed for the use of asymptotic resluts by simulation for different values of P. The asymptotic variances were found to be almost equal to exact and/or simulated variances for moderate sample sizes. These results are given in a technical report by the authors (1976).

REFERENCES

Best, D. J., (1974). "The Variance of the Inverse Binomial Estimator," Biometrika, 385-386.

Chakrabarty, R. P. and Rao, J. N. K., (1967). "The Bias and Stability of Jack-Knife Variance Estimator in Ratio Estimation," Proc. Amer. Statist. Assoc., (Social Statistics section), 326-331.

Chakrabarty, R. P., (1973). "A Note on the Small Sample Theory of the Ratio Estimator in Certain Specified Populations," Jour. Indian Soci. Agri. Statist., Vol. 25, No. 2, 49-57.

Cochran, W. G., (1963). Sampling Techniques, Wiley, New York.

Feller, W., (1968). An Introduction to Probability Theory and Its Applications, Vol 1, 3rd Ed., Wiley, New York.

Haldane, J. B. S., (1945). "On a Method of Estimating Frequencies," Biometrika, 33:222-225.

Kendall, M. G. and Stuart, A., (1968). The Advanced Theory of Statistics, Vol. 1, London, Griffin.

Scheaffer, R. L., (1974). "On Direct Versus Inverse Sampling form Mixed Populations", Biometrics, 30:187-198.

Tsai, P. J. and Chakrabarty, R. P., (1976). "Comparison of Direct and Inverse Sampling for Estimation of Population Proportions," Technical Report, Dept. of Statistics and Computer Science, Univ. of Georgia.

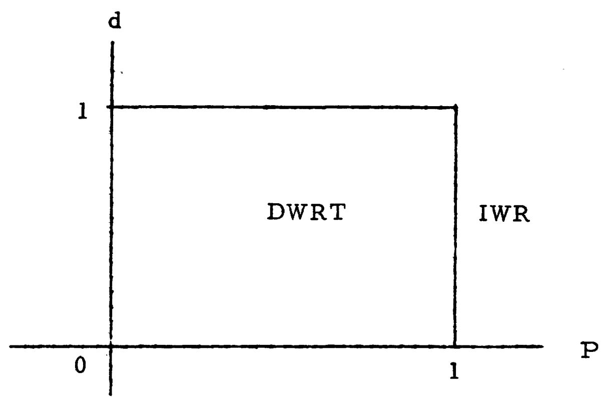


Figure 1: The Region of the Stable Variance Estimators in DWR and DWTR

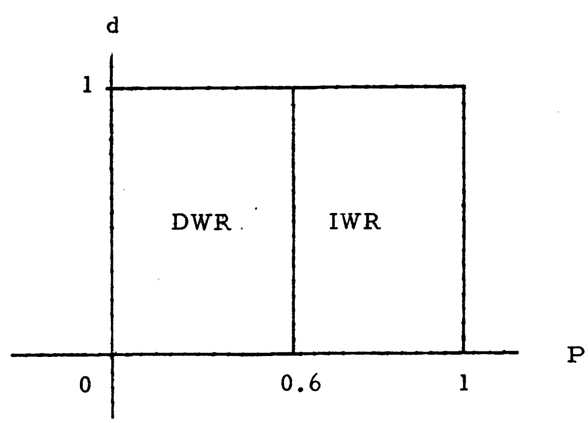


Figure 2: The Region of the Stable Variance Estimators in DWR and IWR

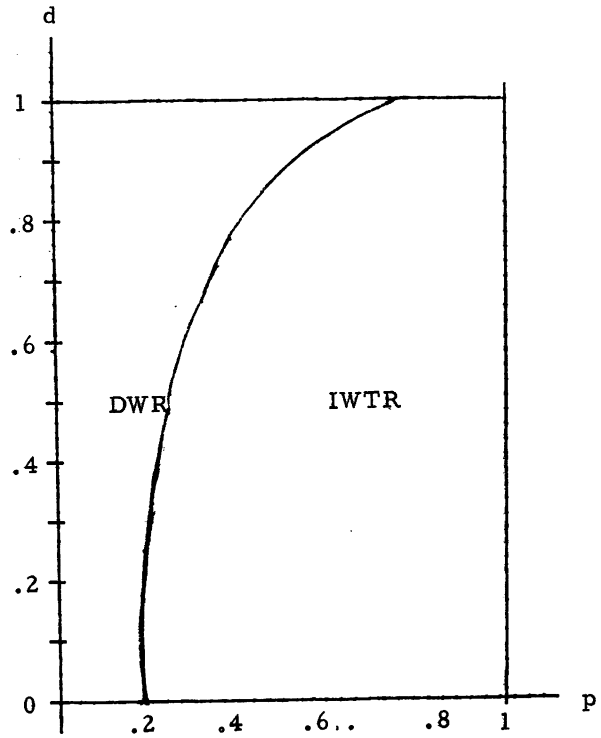


Figure 3: The Region of the Stable Variance Estimator in DWR and IWTR.

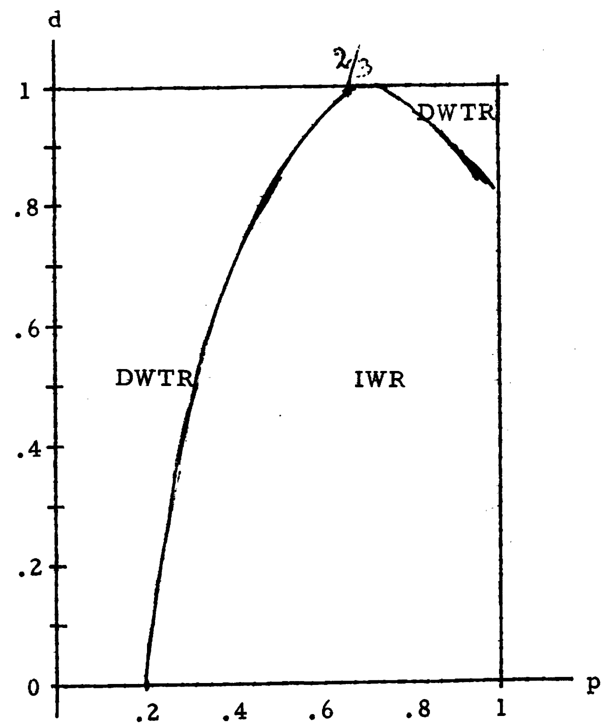


Figure 4: The Region of the Stable Variance Estimator in IWR and DWTR

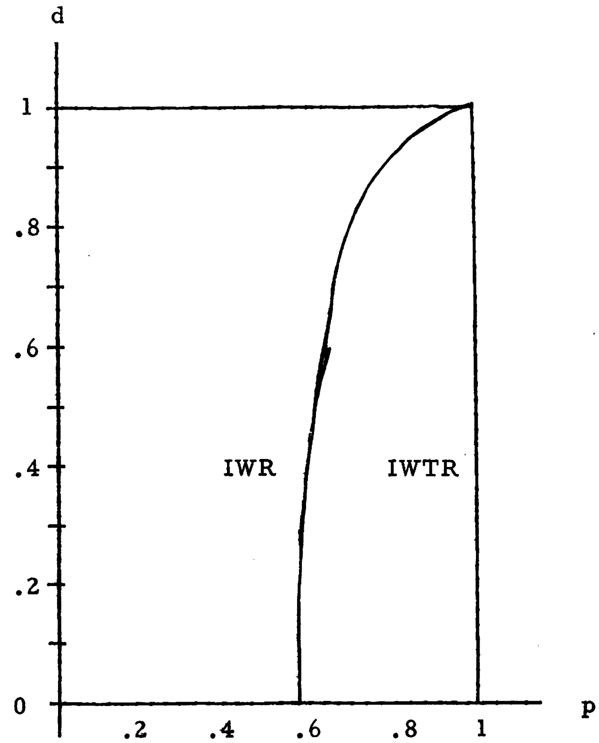


Figure 6: The Region of the Stable Variance Estimator in IWR and IWTR

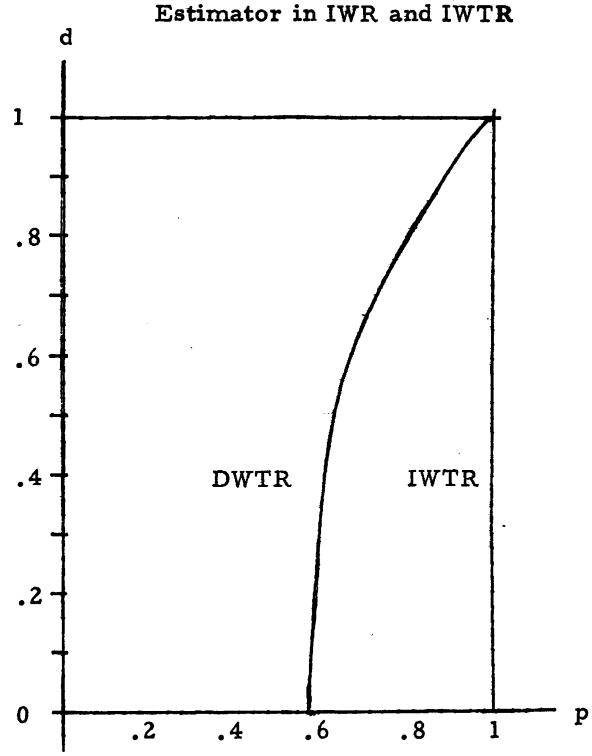


Figure 5: The Region of the Stable Variance Estimator in DWTR and IWTR

SEGMENTED REGRESSION ANALYSIS IN DEALING WITH CROSS-SECTION, QUALITATIVE DATA-- IN APPLICATION TO CONSUMER ATTITUDES TOWARD GASOLINE CONSERVATION MEASURES

Y. C. Chang and K. S. Kim
University of Notre Dame

INTRODUCTION

The purpose of this paper is to illustrate that in dealing with cross-sectional multiple regressions employing qualitative variables measured on a semantic differential scale, appropriate segmentation of the observations into separate regressions generally improves the fit of the regression equations. Within the context of marketing research, the problem of identifying market segments and the variables within each segment that are related to consumer behavior has extensively been discussed in the work by G. D. Hughes (1966) and H. L. Steele (1964). The effectiveness of using the technique of segmented regressions is further illustrated here using the recent survey data on public response to gasoline conservation measures in the United States (see Y. C. Chang and K. S. Kim [1976]).

Illustration

That segmented regression runs improve the fit of the equation for the case of qualitatively differentiable data is illustrated in Figure 1. Suppose that the respondents can be divided into two setments A and B. Let the cluster of observations encircled by points ABCDEF represent segment A; and that of observations by points abcde representing segment B. The equations estimated by least squares method are shown by line KK' and LL' for each respective segment; and by line MM' for the case in which the entire observations are treated as a single, homogeneous segment. This particular example shows that there is a noticeable improvement in the fit of the regression equation as a result of the segmentation of the respondents.

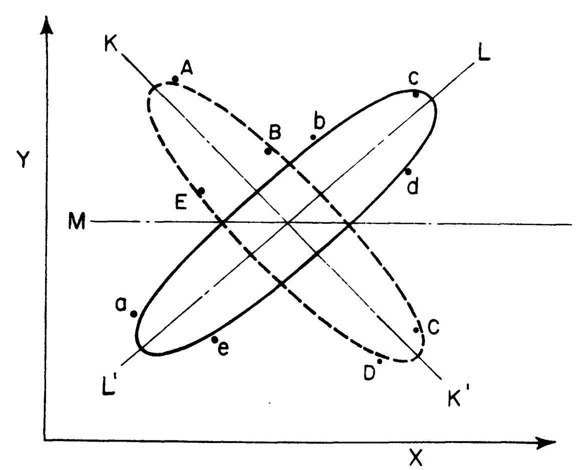


Figure 1. The scatter points and regression lines for the total and segmented population

THE DATA AND VARIABLES

The data analyzed in this study were drawn from a survey conducted jointly by Louis Harris Associates and the Center for the Study of Man in Contemporary Society at the University of Notre Dame. Responses were received from a random sample of 1665 men and women over eighteen years of age living in 100 different locations. The questionnaires included information on the personal backgrounds of the respondents and the usual demographic items. A special effort was made to obtain information on consumer attitudes toward alternative types of involuntary gasoline conservation measures. The particular data used here were obtained in February 1974 at the height of the national energy crisis.

The two questions in the survey relevant to this study were (1) Do you prefer to have a mandatory governmental rationing of gasoline (at the time 35 gallons per week) at current gasoline prices, or do you prefer to have no rationing and pay higher prices for gasoline? The respondents who indicated preferences for higher prices over a rationing system were then asked a second question: (2) How high would the price of gasoline have to go before you would prefer rationing at 35 gallons per week at current prices?

Our interest here is to identify relevant background factors that would explain differences in consumer response relative to the tolerable price level of gasoline. For this the analytical framework is set out simply by postulating a regression equation. The dependent variable is taken as the difference between the tolerable and the then current gasoline price per gallon. Thus the value of the dependent variable in the equation for a respondent opting for the rationing system would be zero.

This price differential is then assumed to be linearly related to such background and locational factors as income, age, occupation, education, sex, race, marital status, urban-rural environment, number of cars owned, as well as to the extent to which automobiles are used. Added to the regression equation are dummy variables for race (nonwhite = 0, white = 1), sex (female = 0, male = 1), urbanity (rural area = 0, urban area = 1), region (east and west coastal areas = 1, other regions =0) occupation (salesman = 1, others = 0), and marital status (single = 0, nonsingle = 1). The availability of public transportation is also included as an explanatory variable. Scores are assigned on the basis of the following four categories determined by the detree of availability: "Available" = 2, "Somewhat available" = 1, "Not available" = 0.

FINDINGS

The first experiment is to fit the regression equation to all the variable available to us using the entire sample. Where information is either missing or inadequate for a particular variable, such data have been assigned the value equal to the sample mean for the variable. The reliability of statistical estimate will not be affected by this procedure because least-squares regression always passes through the sample means of the variables in the equation. The first result shows Multiple R = .294. The F test for the regression, however, shows that it is significant at the .01 level. The low value of R has occurred because of the large amount of random noise in such a large sample in a cross-sectional study. The bulk of the variance explained by the regressions, however, is attributed to only a handful number of the variables. Table 1 reports the results of the first experiment, where the variables for which t values are at least greater than unity are reported.

In the second experiment the entire respondents have been divided into the four different regions in the United States - East, Midwest, South and West. A linear equation is fitted to each of these segments. The results are impressive. In all cases Multiple R's have greatly improved. The proportion of the variance explained by the explanatory variables is, in particular, larger for East and West. The coastal areas were in general more acutely conscious of the energy shortages during the 1974 oil crisis (Louis Harris & Associates Inc. and University of Notre Dame Report, 1974).

Comparisons of the regressions reported in Table 1, and Tables 3 and 4 show an additional interesting fact. Tables 3 and 4 indicate the results of regressions where the respondents are divided into subgroups by sex and the degree of "availability of public transportaiton" in the respondent's residence locality. These two variables have already been included as the dummy variables in the regression equation shown in Table 1. Thus, even when these dummies are included in the regression, the results indicate that Multiple R is uniformly lower in the entire sample case than in the segmented experiments. In theory, this does not have to be the case. The result of segmentation of the respondents would be particularly encouraging if these segmented groups each tend to be more homogeneous with respect to the variables examined.

Finally, variations in the coefficient value and explanatory power of the variables with the further segmentation of the sample are to be noted. Of course, differences in the value and significance of the coefficients identify the characteristics of the behavior of different subgroups. It is interesting to note that in all the cases of segmented regressions where the income and age variables are significant, income is directly related to option for higher gasoline prices while age is inversely related to it.

REGRESSION EQUATIONS INDICATING THE FACTORS DETERMINING THE OPTION FOR HIGHER GASOLINE PRICES

Note: * figures denote estimates significant

TABLE 1

THE ENTIRE SAMPLE

Variable	Reg. Coeff.	F-value
Income	.00005	7.991 *
Age	-.09961	6.877 *
Occupation	.09865	5.447 *
No. of cars owned	-.21767	3.561 *

Multiple R = .294

TABLE 2

REGRESSION EQUATIONS BY REGION

A: EAST

Variable	Reg. Coeff.	F-value
Income	.00006	8.967 *
Age	-.11236	1.006
Occupation	-.02656	.063
No. of cars owned	-.32305	1.406

Multiple R = .486

B: MIDWEST

Variable	Reg. Coeff.	F-value
Income	.00002	1.322
Age	-.07733	1.006
Occupation	.09732	1.572
No. of cars owned	.18632	.635

Multiple R = .377

C: SOUTH

Variable	Reg. Coeff.	F-value
Inc.	-.00000	.003
Age	-.24919	5.962 *
Occupation	.14954	1.427
No. of cars owned	.06737	.004

Multiple R = .343

D: WEST

Variable	Reg. Coeff.	F-value
Income	.00003	.740
Age	-.03729	.103
Occupation	.28496	5.386 *
No. of cars owned	-.08071	.058

Multiple R = .406

TABLE 3

REGRESSION EQUATIONS BY SEX

A: MALE

Variable	Reg. Coeff.	F-value
Income	.00007	3.974 *
Age	-.26030	8.998 *
Occupation	.11411	1.670
No. of cars owned	-.14590	.260

Multiple R = .374

B: FEMALE

Variable	Reg. Coeff.	F-value
Income	.00011	3.285 *
Age	-.15907	1.674
Occupation	.07034	.252
No. of cars owned	-.58154	2.559 *

Multiple R = .370

TABLE 4

REGRESSION EQUATIONS BY
AVAILABILITY OF PUBLIC TRANSPORTATION

A: AVAILABLE

Variable	Reg. Coeff.	F-value
Income	.00011	4.954 *
Age	-.08587	.464
Occupation	-.05296	.149
No. of cars owned	-.32814	.703

Multiple R = .403

B: SOMEWHAT AVAILABLE

Variable	Reg. Coeff.	F-value
Income	.00005	.928
Age	-.20248	3.091 *
Occupation	.10762	.871
No. of cars owned	-.60645	2.369 *

Multiple R - .344

C: NOT AVAILABLE

Variable	Reg. Coeff.	F-value
Income	.00007	1.476
Age	-.39740	8.443 *
Occupation	.24144	2.525 *
No. of cars owned	.05690	.032

Multiple R = .655

REFERENCES

1. Hughes, G. D. 1966 "Developing Marketing Strategy Through Multiple Regression," Journal of Marketing Research, 3, 412-415.

2. Steele, H. L. 1964 "On the Validity of Projective Questions," Journal of Marketing Research, 1, 46-49.

3. Louis Harris Associates Inc., and University of Notre Dame, 1974. Public Response to Gasoline Shortage, Report.

4. Chang, Y. C. & K. S. Kim, 1976 "The Socio-Economic Determinants of Gasoline Conservation Measures: Rationing or Higher Price?" MIDWEST AIDS Proceedings, pp. 423-426.

A SURVEY OF NONRESPONSE IMPUTATION PROCEDURES [1]

David W. Chapman, Westat, Inc.

1. Introduction - A good sampling plan for a sample survey will include an extensive effort, whenever necessary, to obtain a usable response for each unit selected into the sample. Various aspects of the design, such as clustering and allocation of resources, are adapted to make this feasible and practical. However, in spite of such efforts there will always be some nonresponse in large-scale surveys. Furthermore, as indicated in a report on methodology for the Current Population Survey (12,p.53), there are no known unbiased or even consistent methods of imputing for nonresponse (unless special assumptions are made regarding the nature of the nonresponse.)

Rather than imputing for nonresponse at the time survey tabulations are prepared, tabulations could be presented with the amounts of nonresponse reported in a nonresponse category. This would allow users of the data to select their own method of making nonresponse imputations. However, the additional burden of having to compute nonresponse adjustments may not be worth having the choice of imputation method. Furthermore, users would have to make imputations from the tabulated data, and some of the related information available at the tabulation stage could not be used in these imputation procedures. It therefore appears to be more appropriate to make imputations at the time tabulations are prepared, thus eliminating the nonresponse category from the tables (except perhaps to allow for item nonresponse). If nonresponse adjustments are made, the level of nonresponse should always be reported when presenting survey results.

Pritzker, Ogus, and Hansen (10,p.445) indicate that, based on extensive experience, if the survey nonresponse rate is less than five percent, any plausible method of nonresponse imputation will probably provide acceptable results. However, in many sample surveys, nonresponse rates are substantially higher than five percent. Even with interviewer surveys which include several call-backs to households and telephone followup efforts, nonresponse rates will sometimes equal or exceed 20 percent.

In such cases, the method of nonresponse imputation can have a substantial effect on the values and biases of the survey estimates. Much research has been carried out in an attempt to discover imputation methods which reduce or minimize the nonresponse bias.

In this paper an attempt will be made to summarize some of the procedures which have been used. These procedures will be discussed in two sections: one dealing with item nonresponse, and the other dealing with total (questionnaire) nonresponse.

2. Imputation for Item Nonresponse - In most surveys some of the respondents refuse, neglect, or are unable to complete one or more questionnaire items even though they do complete most of the items. (For example, income is sometimes not supplied in a household survey.) Item nonresponse also arises when a response is received which, on the basis of the editing procedures, is determined to be unacceptable. For respondents with missing items, the information available from the completed items can be used to help impute responses for the incomplete items. In fact, sometimes there is redundant information in the questionnaire and a missing item can be inferred appropriately from other responses. In other instances, an approximate value for a missing item may be obtained by considering the general relationship between this item and another item. For example, for purchases of homes, closing costs might be estimated as a percentage of the price of the home.

Two procedures which have been used by the Bureau of Census (among others) to impute for missing values or presumably incorrect values are the "cold-deck" and "hot-deck" methods. These two procedures will be discussed in this section.

2.1 The Cold-Deck Procedure for Imputation[2] - Basically, the cold-deck procedure uses values from some prior distribution to substitute for missing responses. The distribution used is usually obtained from a previous survey taken from essentially the same population. For example, for the Census an appropriate distribution would be obtained from the previous Census, or a recent household survey.

To use a distribution of prior responses for imputation, the responses are classified by one variable or jointly by two or more variables that are reported. An attempt is made to define cross-categories (or cells) in such a way that responses will be relatively homogeneous within cells and heterogeneous between cells. There must be at least one response in each cell available for imputation. The responses in each cell are stored in the memory of the computer. A cold-deck distribution is prepared in advance for each cell.

For each missing item for a particular respondent to the current survey, the values of the appropriate completed items are noted to identify the relevant cell. The respondent is associated with the cell corresponding to the values of the items. A value is then selected from the responses in the cold deck included in the same cell. This value is usually selected at random or systematically.

As an example, suppose that the age of a respondent could be placed in a cell determined by sex and household relationship, and perhaps by the age of another member of the household. Then the age of one of the cold-deck respondents selected at random from the same cell would be inserted for the missing age.

2.2 The Hot-Deck Procedure for Imputation [3] - An objection to the cold-deck procedure is that it does not utilize data obtained from the current survey. The hot-deck procedure does use the current responses to substitute for missing items.

As with the cold-deck procedure, crossclassifications (or cells) are identified by one or more relevant variables. Initial values for each cell must be supplied from a cold deck to initiate the procedure. Then new responses are supplied for each cell from the new (or hot) deck as they appear in a pass through the file. The file may be arranged in order based on relevant variables before the procedure begins. A response remains in a cell until another respondent appears who has the same characteristicts (i.e., is in the same cell) and has a response for the particular item.

Whenever an item is missing for a respondent, he is first identified with the appropriate cell, based on the responses he does supply. Then the value retained in that cell is imputed to the respondent with the missing value. As an example, suppose that age, sex, race, household relationship, and level of education were used to define cells for imputing income values. A respondent whose income is not provided is placed into the appropriate cell as determined by his responses to the above items. The value of income in that cell (ie., the income of the respondent having the same characteristics and appearing most recently in the file sequence) would be taken as the missing income. In order to avoid using the same income value repeatedly, several income values could be stored in a cell and these values could be used in rotation, if necessary.

The method described above of imputing the preceding value for a missing item is better than using a random selection from all those in the sample falling into the same cell. This is because there is usually some special ordering of the respondents which indicates that an imputed value from a respondent close in the file would be better than one picked at random from the cell. Also, it is a more convenient procedure for computer processing.

There are some possible variations on the use of the hot-deck procedure. One would be to use as an imputed value one obtained from a regression of the particular item on several of the other items. A regression equation could be developed from either a hot deck or a cold deck.

Another variation would be to use a moving average of values in a cell to substitute for a missing value. This procedure would prevent extreme values from being duplicated and would therefore reduce slightly the variances of the estimates. However, if the ordering of the respondents were important, such a procedure would contain slightly more nonresponse bias.

3. Imputation for Total Questionnaire) Nonresponse - The hot-deck procedure described above for imputing missing items could also be used to impute values for an entire questionnaire to survey nonrespondents. As described by Pritzker, Ogus and Hansen (10,p.460), this was done in the 1960 Census by substituting for a nonresponding household the questionnaire responses of the previously listed responding household. This procedure amounts to doubling the weight[4] of the respondents whose records are duplicated. Such a procedure can yield somewhat larger variances of the survey estimates than would the procedure of weight adjustment discussed below. Hansen, Hurwitz, and Madow (3,pp.232-233) show that the maximum increase in variance is about 12 percent for the method of duplicating records.

In many surveys imputation for nonresponse is carried out by adjusting the weights of the respondents in some way to account for the nonrespondents. Alternate methods of making weight adjustments plus other methods of imputation for survey respondents will be discussed in the following sections.

3.1 The Use of a Single Weight Adjustment to Account for Nonresponse - The simplest type of nonresponse adjustment is to make one overall weight adjustment. This adjustment would be equal to the sum of the initial weights of all units selected into the sample divided by the sum of the weights of the respondents. Such an adjustment "weights up" the respondents to the total sample. (If all units selected have the same initial weights, this adjustment would equal the sample size divided by the number of respondents.)

The nonresponse bias associated with this procedure can be derived in a simple case. Suppose that a simple random sample of n units is selected from the N units in the population. The basic sampling weight for each unit selected is N/n (ie, the inverse of the selection probability). Let n_1 represent the number of the n sample units that respond to the survey. If one overall weight adjustment is used in this case, it would be n/n_1 since all sample units have the same basic weight. Also, assuming no other weight adjustments are used, each of the n_1 respondents would have the same final weight, $(N/n)(n/n_1)$.

The basic formula for estimating a population mean from weighted data is the following:

$$\bar{x} = \sum_{j=1}^{n_1} w_j x_j \Big/ \sum_{j=1}^{n_1} w_j \qquad (1)$$

where

n_1 = the number of respondents,

w_j = the final weight assigned to the jth respondent,

x_j = the value of the variable (item) for the jth respondent.

In this case, since the weights of the respondents are all equal, the estimated mean in equation 1 reduces to a simple unweighted mean of the n_1 respondents.

In this case the expected value of $\bar{x}$ is equal to $\bar{X}_{.1}$, the mean of the variable for all those in the population who would respond if selected for the survey.

The bias of $\bar{x}$ for this case can be written as follows:

$$\text{bias}(\bar{x}) = E(\bar{x}) - \bar{X} = \bar{X}_{\cdot 1} - [R\bar{X}_{\cdot 1} + (1-R)\bar{X}_{\cdot 2}]$$
$$= (1 - R)(\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}) \qquad (2)$$

where

R = the population response rate (ie, the proportion of the N population units that would respond if selected for the survey),

$\bar{X}_{\cdot 2}$ = The mean of the variable for all those in the population who would not respond if selected.

As expected, the bias of $\bar{x}$ depends on two factors: (1) the population nonresponse rate, $1 - R$, which is a function of the data collection procedures; and (2) the difference between the population mean for respondents and the mean for nonrespondents, $\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}$.

In an attempt to reduce the nonresponse bias of this simple adjustment procedure, weighting classes are often defined based on the characteristics available for both respondents and nonrespondents. Separate nonresponse weighting adjustments are made within each weighting class. This procedure is discussed in the next section.

3.2 The Use of Weighting Classes to Make Nonresponse Adjustments - Suppose that the population is partitioned into c classes, based on the values of one or more survey items. Let $P_1, P_2, \ldots P_c$ represent the proportions of the population members contained in each of these classes. Also, let $R_1, R_2, \ldots R_c$ be the proportions of the units in these weighting classes that would respond if selected for the survey.

As in the previous case, suppose that a simple random sample of n units is selected from the N population units. Let $n_1, n_2, \ldots n_c$ be the number of sampling units falling into each of the classes. Of course, the n_i values are random variables and their sum must equal n. Also, let $n_{11}, n_{21}, n_{31}, \ldots, n_{c1}$ represent the number of survey respondents in the c classes. The basic sampling weight (ie, inverse of the selection probability) would be (N/n) for each sample unit (as in the previous case). However, the nonresponse adjustments would vary from class to class. For each respondent in the ith class this adjustment would equal (n_i/n_{i1}), which is the sum of the sampling weights of all sampling units falling into the ith cell divided by the sum of the sampling weights of all respondents falling into the ith cell.

The estimate, $\bar{x}_1$, of the mean would then be computed as

$$\bar{x}_1 = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n_{i1}} (N/n)(n_i/n_{i1}) x_{ij}}{\sum_{i=1}^{c} \sum_{j=1}^{n_{i1}} (N/n)(n_i/n_{i1})} = \sum_{i=1}^{c} p_i \bar{x}_{i1}, \qquad (3)$$

where

$\bar{x}_{i1}$ = the sample mean among respondents in the ith weighting class,

p_i = the proportion of the sample falling into the ith weighting class.

The expected value of $\bar{x}_1$ is the following:

$$E(\bar{x}_1) = \sum_{i=1}^{c} P_i \bar{X}_{i1} \qquad (4)$$

where

$\bar{X}_{i1}$ = the mean of the variable for all those in the population contained in the ith weighting class who would respond if selected for the survey.

The bias of $\bar{x}_1$ can be written as follows:

$$\text{bias}(\bar{x}_1) = \sum_{i=1}^{c} P_i(1-R_i)(\bar{X}_{i1} - \bar{X}_{i2}). \qquad (5)$$

It is useful to compare the bias of $\bar{x}_1$ given in equation 5 to that of $\bar{x}$ given in equation 2. If for each of the c weighting classes $(\bar{X}_{i1} - \bar{X}_{i2})$ equals $\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}$, the biases of $\bar{x}$ and $\bar{x}_1$ are identical. Also, the bias of $\bar{x}_1$ is equal to that of $\bar{x}$ if all the class response rates, $R_1, R_2, \ldots, R_c$, are equal to the overall nonresponse rate, R.

However, if the $\bar{X}_{i1} - \bar{X}_{i2}$ values tend to be less (in absolute value) than $\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}$ and the response rates (ie, the R_i values) vary from class to class, the nonresponse bias will be reduced by the use of the weighting classes to make nonresponse adjustments. Therefore, the successful application of this procedure requires the identification of survey characteristics which will define weighting classes which vary both with respect to response rates and survey estimates. Furthermore, the characteristics used to define weighting classes must be available for both the respondents and nonrespondents. This requirement will, in many surveys, severely limit the choices of variables to use to define weighting classes.

There are many surveys in which the procedure discussed above is used to impute for nonrespondents. Among them are the Health Examination Survey (8,p.6) and the Current Population Survey (12,p.53). In Cycle I of the Health Examination Survey, seven age-sex weighting classes were defined within each of 42 primary sampling units (PSU's) for a total of 294 separate cells. Nearly half of the 294 nonresponse adjustments were between 1 and 1.10 and the three largest estimates were between 2.01 and 2.10. In the CPS, the PSU's are grouped together based on the population and labor-force characteristics of the strata from which the PSU's were selected. Within groups of PSU's respondents are placed in six cells based on race-residence characteristics.

In some cases the total number of members in each weighting class is known (or a good estimate is available) and used in the nonresponse adjustment. In such cases, the weights of respondents in a cell are weighted up to the "known" total. This procedure is closely related to stratification after sampling, discussed by Hansen, Hurwitz and Madow (3,p.232; 4, pp. 138-139). The bias of the estimate of the mean using this procedure is the same as that given in equation 5 for $\bar{x}_1$, assuming the population totals are known exactly. However, the variance and therefore the mean square error would be less for the procedure based on known totals.

Care must be taken in the application of these two imputation procedures. As demonstrated in Hansen, Hurwitz and Madow (4,pp.138-139), the variance of the estimated mean can be increased by weighting up to cell totals if the number of respondents in the cells is small. As a rule of thumb, a minimum of 20 respondents is used for the weighting cells in the CPS (12,p.53). Furthermore, for the CPS a maximum of 2.0 is taken for the nonresponse adjustment factor. In cases in which the adjustment exceeds 2, cells are combined to the extent necessary to reduce the adjustment to 2.0 or less.[5] In the Health Examination Survey (8,p.6), the 294 weighting cells average about 25 respondents each.

The choices of which variables to use to define weighting classes are usually based on which variables have the higher correlations with the zero-one response variable and with the characteristics for which survey estimates are made. It is assumed that variables which show a high correlation among respondents for survey characteristics to be estimated would show high correlations among survey nonrespondents. If so, then the use of such variables to define weighting classes would presumably minimize the nonresponse bias. The decisions on priorities of the use of variables to define weighting classes are largely subjective. These decisions involve choices as to which variables to collapse whenever weighting cells have to be combined to provide adequate numbers of respondents per cell.

A procedure which can be used to determine weighting classes objectively from a pool of possible weighting variables is the AID programmed procedure. One way this can be done is discussed in the next subsection.

3.3 <u>The Use of the AID Programmed Procedure to Define Weighting Classes</u> - The AID programmed procedure can be used to select which variables to use in weighting classes and also to specify which crossclassifications of these variables should be used to define weighting classes.[6] Using this procedure the sample would be divided sequentially into subgroups in a way to maximize the amount of variability explained in some dependent variable. The dependent variable used could be the zero-one response variable, or a survey questionnaire item. As a first step, the sample would be split in half based on the categories of a single variable. The variable selected from the pool of variables is the one which provides for the maximum amount of explained variance by a division into two groups. Next, one of these two groups is split again in such a way as to maximize the explained variance in the dependent variable. This procedure of defining new subgroups to account for the maximum amount of variance is repeated until the weighting classes become as small as is allowed in the specifications, or until it is no longer possible to explain meaningful proportions of remaining variance.

There has been very little investigation of the use of AID in this capacity. In a report prepared for NCHS by Chapman (2,pp.10-20), the use of AID was tested on data collected in the Health and Nutrition Examination Survey. The basic conclusion from this investigation was that the use of AID to define specific weighting classes for nonresponse adjustments does not appear to be feasible. The specific classes identified by AID can be very complex and would probably be rather awkward to work with in practice. Also, some of the classes contained a very small number of respondents, which can increase the variance (as discussed earlier). Finally, there is no easy way of merging an AID analysis based on the zero-one response variable as the dependent variable with that based on one or more survey items as dependent variables.

Perhaps the most useful information from the AID results is obtained by noting which independent variables are used most often in defining "optimal" splits in the sample subgroups. These independent variables would probably be most useful in defining weighting classes of the type discussed earlier in Section 3.2.

3.4 <u>The "Raking" or "Balancing" Procedure for Nonresponse Imputation</u>[7] - The "raking" procedure is one which allows the use of a large number of variables to define weighting classes simultaneously, without being concerned about the number of respondents in crossclassifications.

This method utilizes known marginal totals for the categories of two or more characteristics selected for weighting variables. These characteristics must, as before, be known for nonrespondents as well as respondents. First, the weights of the survey respondents are blown up to the given marginal totals for one of the variables. Next, the weights of the respondents, as adjusted in the prior step, are further adjusted to add to the given marginal totals for one of the other variables. This procedure is repeated for each of the variables used for the raking procedure. At this point, only the last variable dealt with will be sure to have desired marginal weight totals. However, the procedure can be repeated and the marginal totals converge to the desired numbers for all variables. The convergence proof is due to Ireland and S. Kullback(5).

The resulting adjustment applied to a particular respondent is the product of the adjustments made for the marginal total for each variable for each iteration. Estimation based on these weights has a justification in statistical information theory.[8]

As an example of the raking procedure, suppose that two variables are used in the adjustment process. Let the given marginal total for the ith category of the row variable be denoted as $N_{i.}$, and let the known total for the jth category of the column variable be noted as $N_{.j}$. Also, let $n_{i.}$, $n_{.j}$, and n_{ij} represent marginal sample size for the ith category of the row variable, the marginal sample size for the jth category of the column variable, and the sample size for the ijth cell. (These sample sizes can be taken to be sums of respondent weights.) Then, adjusting the row totals first, the new frequency (or sum of weights) of the respondents in the ijth cell is the following:

$$N_{ij}^{(1)} = (N_{i.} / n_{i.})n_{ij} \quad (6)$$

Next, the cell frequency $N_{ij}^{(1)}$ is replaced by the following value:

$$N_{ij}^{(2)} = N_{ij}^{(1)} (N_{.j} / N_{.j}^{(1)}) \quad (7)$$

where

$N_{.j}^{(1)}$ = the marginal total for the jth column after the first adjustment is made using equation 6.

Repeated iterations of this process can be made until the desired level of convergence on the marginal totals is reached.

3.5 The Use of Regression in Weighting Adjustments - A procedure using multiple regression in nonresponse imputation has been used by Astin and Molm (1) for a follow-up survey of college freshmen. Basically, the zero-one response variable is regressed on some set of independent variables which are available for both respondents and nonrespondents. The value of the regression equation for each respondent is the estimated response rate or probability of responding for population members with the same values of the independent variables. The nonresponse weight adjustment for each respondent is taken as the inverse of the value of the regression equation.

In the application of this technique the multiple correlation coefficient was less than .25. Since this indicates that only about six percent of the variation in the zero-one response variable was explained by the regression equation, there is some doubt regarding the use of this procedure. However, to compare this procedure with the weighting-class-type procedure, corresponding measures of the explained variance for the zero-one response variable would have to be observed. The low proportion of explained variation may be a result of the linearity assumption underlying the regression model which was used. The implications of the linearity assumptions are discussed for a simple case by Chapman (2,pp.60-61). A nonlinear regression model may lead to higher proportions of explained variation in the zero-one response variable. If this is the case, this method of nonresponse imputation may be more appropriate for nonlinear regression.

There are other ways that regression could be used in imputation. For example, each survey item could be regressed on the variables that are available for both respondents and nonrespondents. Of course, estimates of the regression coefficients would have to come from the respondent sample. Imputed values for the questionnaire items would be obtained for a nonrespondent from the regression equations.

A difficulty with this procedure would be the need for a large number of regression equations -- one for each questionnaire item. Another problem would be the limited information that is available for nonrespondents. That is, there may not be enough meaningful independent variables available for the regressions to be worthwhile. Perhaps this procedure would be more useful in the case of imputation for item nonresponse since a larger number of independent variables would be available in that situation.

3.6 The Use in Imputation of the Amount of Effort Needed to Obtain Response - If whether or not an individual selected for a sample survey participates in the survey is correlated to the measurements taken, then it seems plausible that the number of calls required to obtain participation would also be correlated to the measurements taken. If so, the number of calls required per respondent could be useful in nonresponse imputation.

One way that the number of calls could be used would be to make nonresponse weight adjustments among only those respondents who agreed to participate after several calls. Weighting classes would be defined among the nonrespondents and the "late cooperators". The late cooperators would receive weight adjustments computed in a way similar to those discussed in Section 3.2.

This procedure would minimize the bias if, indeed, the survey characteristics of the nonrespondents were more alike those of the late cooperators than those for all survey respondents. However, the validity of this assumption is questionable. It might hold for some surveys and not for others. With regard to the CPS, Waksberg and Pearl (14,p.232) indicate that there is no support for the hypothesis that the characteristics of the nonrespondents become more like the respondents as the number of visits required for interview increases. This statement was based on results from an intensive follow-up of CPS nonrespondents in which about 40 percent of the original nonrespondents were interviewed.

This procedure can have an undesirable effect on the variances of survey estimates. If the number of late cooperators is not considerably larger than the number of nonrespondents, the nonresponse weight adjustments could be relatively large. If so, the variances of survey estimates would be increased substantially.

In the imputation process another method of using the number of calls needed to complete the interview would be to try to project a mean response for nonrespondents. That is, for a particular survey item, a mean response would first be computed among respondents requiring only one call. The corresponding mean would also be computed among those requiring two calls, among those requiring three calls, etc. If the mean responses were plotted against the number of calls, a trend might be apparent.

This procedure was investigated by Chapman (2,pp.51-59) for data collected in the Health and Nutrition Examination Survey. In this case there were many different patterns observed for the various survey items. It was not possible to determine a general trend. Also for most items, the trend of mean response as a function of the number of calls was not evident enough to even attempt to project a mean value for the nonrespondents. Even for the few items for which the trend was apparent, the method appropriate to extrapolate to the nonrespondents is unclear. Consequently, the use of degree of persuasion in the imputation process for that survey did not appear to be feasible.

3.7 Imputation by Substitution of Additional Selections from the Population - For surveys in which it is feasible, imputations for nonrespondents are sometimes made by selecting substitute units from the population to take the place of the nonrespondents. For such cases an attempt is made to obtain a substitute with characteristics which are similar to those of the nonrespondent. This may be done by selecting an additional sampling unit at random from the same stratum or cluster as the unit which did not respond, or may involve a substitute picked on a subjective basis to appropriately "represent" the nonrespondent, such as a neighbor. When such a substitute sampling unit is obtained for the sample, the substitute unit is weighted as though it had been initially selected.

A possible difficulty with a substitution procedure of this type is that the effort put forth to obtain a response from each of the originally selected sampling units may not be as strong as it would have been if no substitution procedure were used. This is a serious problem since the only satisfactory way to deal with nonresponse is to keep it to a low level. Therefore every effort should be made to obtain usable responses from those units originally selected before substitutions are made.

Also, when substitutes are used, it is important to keep in mind that the total sample (ie., original respondents plus substitutes) is not equivalent to a probability sample of the same total size from the population. This is because of the bias introduced due to the use of substitutes in place of some of the originally selected units. Therefore, when this procedure is used, the amount of substitution involved should be reported.

If the above problems are taken into account and kept under reasonable control, then the substitution procedure is good if adequate substitutions are made. In particular, it has the advantage over the weighting-class method in that it does not involve any inflation of weights which causes some increase in the variances of the estimates. Also, it does provide more respondent data than the other procedures.

Of course, it is usually not possible to obtain a substitute for each nonrespondent. Therefore, even when substitution is used, one of the previous methods of adjusting weights must also be used to some extent.

As an example, Westat Research was a subcontractor to the Educational Testing Service to help design a sample of 448 elementary schools in which to administer achievement tests. The test design required exactly 448 schools. Therefore substitutes had to be obtained for each of the nonresponding schools.

The first level substitute for a nonresponding school was taken to be that school, if any, located in the same district, having the same grade structure, and having similar levels of enrollment, mean income of the surrounding community, and percent minority of the students. If such a substitute was not available, no other school in the same district was allowed as a substitute. For the second, third, fourth and fifth priority level substitutes, schools were selected at random from the same stratum as the nonresponding school.

As a result of a superb effort on the part of ETS personnel (Western Office), all 448 slots were presumably filled. Unfortunately, improper test administration forced three of these schools out of the respondent group, leaving a total of 445 responding schools. Weight adjustments using classes defined by strata were used to impute for the three missing schools.

FOOTNOTES

[1]I would like to express my sincere appreciation to Morris H. Hansen and Sidney A. Jaffe for their many helpful suggestions regarding the content of this paper. They also read over the first draft and made many useful comments.

[2]The discussion of the cold-deck imputation procedure given here is based on a description of the procedure given by Svein Nordbotten (9,pp. 26-27).

[3]The discussion of the hot-deck imputation procedure is based primarily on descriptions by Svein Nordbotten (9,pp.28-29) and the Bureau of the Census (13,pp.22-23).

[4]The weight of a respondent is a quantity which is used to give the respondent his appropriate representation in the calculation of the survey estimates. This weight consists of the product of (1) the inverse of the selection probability, (2) any ratio adjustments to known totals, and (3) nonresponse adjustments.

[5]This use of 2.0 as a maximum weight is based, to a large extent, on the overall CPS nonresponse rate of only 5 percent. In surveys with higher nonresponse rates, the maximum adjustment allowed is probably higher.

[6]A detailed description of the AID programmed procedure is given by Morgan and Sonquist (7).

[7]The general description of the raking procedure given here is based on a description by Rosenblatt (11, especially pp.4-6).

[8]This is discussed by Rosenblatt (11,p.5) and is covered in detail by Kullback (6).

REFERENCES:

1. Astin, Alexander W. and Linda D. Molm, Correcting for Nonresponse Bias in Follow-up Surveys, 1972. An unpublished manuscript available from the Office of Research of the American Council on Education, 1 Dupont Circle, Washington, D. C.
2. Chapman, David W., "An Investigation of Nonresponse Imputation Procedures for the Health and Nutrition Examination Survey," prepared for the Division of Health Examination Statistics, National Center for Health Statistics, HEW, by Westat, Inc., 1974.
3. Hansen, M.H., W.N. Hurwitz, and W.G. Madow, Sample Survey Methods and Theory, Vol. I,

New York: John Wiley and Sons, Inc., 1953.

4. Hansen, M.H., W.N. Hurwitz, and W.G. Madow, Sample Survey Methods and Theory, Vol. II, New York: John Wiley and Sons, Inc., 1953.
5. Ireland and Kullback, S., "Contingency Tables with Given Marginals," Biometrika, Vol. 55, No. 1 (March 1968), pp.179-188.
6. Kullback, S., Information Theory and Statistics, New York: John Wiley and Sons, 1959; Dover Press, 1968.
7. Morgan, James N. and John A. Sonquist, "Problems in the Analysis of Survey Data, and a Proposal", Journal of the American Statistical Association, 58 (June 1963), pp.415-435.
8. National Center for Health Statistics: Cycle I of the Health Examination Survey: Sample and Response, United States, 1960-1962. Vital and Health Statistics. PHS Pub. No. 1000 - Series 11 - No. 1 Public Health Service. Washington: U. S. Govt. Printing Office, April 1964.
9. Nordbotten, Svein, "Automatic Editing of Individual Statistical Observations", Conference on European Statisticians, Statistical Studies, No. 2, United Nations, New York, 1963.
10. Pritzker, L., J. Ogus and M.H. Hansen, "Computer Editing Methods - Some Applications and Results", Bulletin of the International Statistical Institute Preceedings of the 35th Session, Belgrade 41 (September 1965), pp. 417-441.
11. Rosenblatt, Harry M., "Study of Proposed Ratio Estimators for 1970 Census". Prepared for Census History, Statistical Research Division of the U. S. Bureau of the Census, March 11, 1971.
12. U. S. Bureau of the Census, The Current Population Survey - A Report on Methodology. Technical Paper No. 7, U. S. Government Printing Office, Washington, D. C., 1963.
13. U. S. Bureau of the Census, United States Censuses of Population and Housing, 1960: Processing the Data. Washington, D. C.,1962.
14. Waksberg, Joseph and Robert Pearl, "New Methodological Research on Labor Force Measurements". Proceedings of the American Statistical Association, Social Statistics Section, 1965, pp. 227-237.

ON UNBIASED ESTIMATION IN THE AREA AND LIST FRAMES

Nicholas J. Ciancio and Robert D. Tortora, Statistical Reporting Service, USDA

INTRODUCTION

In a standard sampling text one defines a target population, specifies a sampling method, then caculates an estimate and variance. From a practical standpoint one must: (a) define the target population, (b) determine sampling frame units to reach the target population, and (c) define rules to associate the target population units with the sampling frame units.

Since the sampling units are the units that the sampling frame is divided into for sampling purposes they may not be the same as the target population of units. This affects (i) the sample design in that (a), (b), and (c) above have some impact on whether single or multiple frames are used and how samples are allocated between and within frames, (ii) the questionnaire design since the questions must determine which population unit is being associated with the sampling unit, (iii) the collection procedures since the sampling unit determines what is to be sampled, and (iv) the estimation procedures because they are directly related to the sample design, and should be unbiased if so desired.

Expected values are taken to determine if the estimators are unbiased. These estimators must remain unbiased under the survey operational conditions. This paper will show that under practical situations the estimates remain unbiased when different rules are used to associate the sampling unit and population unit under a simple random sample design.

This problem is discussed for two reasons. First, the Statistical Reporting Service, in its continued attempts to improve survey methodology, evaluates and compares rules to associate target population units with sampling frame units. Secondly, the problem of correctly using the frame to reflect the target population is critical to the success of any survey. The importance of this problem is noted by Cochran [1] and Deming [2] as discussants to a paper presented by Hansen, Hurwitz and Jabine, [3]. However, in [3] no proofs are presented for the surveys they mention. Therefore, this paper demonstrates that such proofs are feasible and that they are necessary to insure that good survey procedures are followed.

TERMINOLOGY

Before attempting to show that unbiased estimates are obtained, we must define the target population units, sampling frames, and describe the rules of association between the target population units and sampling frame units.

The target population is all farms in the 48 states. From [6] a farm consists of the area or areas of land under one operation or management including land owned and rented minus land rented to others on which there will be crops, livestock, poultry, or expected sales of agricultural products at some time during the calendar year.

In the target population there are different types of land operations. One is an individual land operation in which a single person is solely responsible for making management decisions for his business. A joint land operation is one operated by 2 or more persons, each of whom contributes some or all of the money, property, materials or labor to carry on a joint business. Each person participates in the management decisions and shares the profits or losses. Examples of joint arrangements are partnerships, corporations, and institutions or cooperatives. Finally, managed land is an operation whereby a person is paid to make the day-to-day decisions for the farm. The target population has been structured in two ways with regard to joint operation which is discussed under Rules 1 and 2.

There are two frames or partial frames, the area frame and the list frame, that when combined cover the target population used for multiple frame estimation. The area frame consists of all land area within the states. The area frame covers 100% of the population, therefore, the area frame is a complete frame. The land area is classified (stratified) according to land use in order to achieve homogeneity within strata. For the area frame the sampling unit is a small section of land called a segment. A segment is a piece of land with boundaries delineated on a map. Every parcel of land within a segment must be accounted for in the survey.

Within each segment sampled, all farms whose headquarters are within the segment boundaries, are interviewed. Every population unit (farm) is assigned to only one sampling unit (segment) even though pieces of land area associated with the population unit fall within many sampling units since each farm can have only one headquarters. Each sampling unit may contain more than one population unit or no population units.

Situations arise where it will be necessary to distinguish to which sampling unit a farm belongs. This is done by an approach which requires a 1-1 correspondence between farm operators and farms. The approach is needed because it is possible for more than one person to be accepted as the farm operator of a particular farm. For example, suppose two brothers operate a farm jointly and live in different houses. Unless proper rules are formulated this farm could easily have a chance of being sampled twice.

For individual operations the residence of the operator is usually defined as the headquarters. The following are examples of possible rules to help determine the operator of types of jointly operated farms.

(a) In an individual operation only the individual can be associated with the operation. For example, suppose Bob Smith is on the list and Bob Smith is an individual operator then if his name is selected he will report for the farm.

(b) In a joint operation there are three possible kinds of sampling units:

1. A joint operation name is on the list but none of the respective names of individuals

who comprise the joint operation are on the list. Therefore, if Smith Brothers is on the list and that sampling frame unit is selected, the farm operated for the Smith Brothers will be associated with the sampling unit.

2. The joint operations name is not on the list but at least one of the individuals who comprise the joint operation is on the list, then information concerning the joint operation will be reported by the individuals. If Sam Smith is on the list and is a partner in Smith Bros., then if he is selected he will report for Smith Bros. farm. If all partners report for Smith Bros. farm, duplication will result.

3. Both the joint operation name and the names of the individuals who comprise the joint operation are on the list. For example, suppose Smith Bros., Sam Smith and Bill Smith are on the list. Suppose Smith Bros. is comprised of Sam and Bill Smith. If the name Smith Bros. is selected it will report for Smith Bros. If either the name Sam Smith or Bill Smith is selected they will report for individual operations of their own, if any, but not for Smith Bros.

As can be seen the rules given above are not the only rules which can be used. Other rules of association between the sampling unit and population unit can be developed.

(i) When all partners live on the farm, the person who makes most of the decisions should be considered the operator,

(ii) If one partner lives on the farm, and the others live elsewhere, the one living on the farm should be considered the operator,

(iii) If all partners live on the farm, and appear to share equally in the management, the oldest should be considered the operator,

(iv) If none of the partners live on the farm, the oldest should be considered the operator,

(v) In father-son arrangements, accept the definitions of the respondent as to whether it is (1) a partnership, (2) two separate operations, or, (3) one operation with the father in charge and the 4-H or F.F.A. projects of the son merely a part of or incidental to the fathers' overall farming operations.

For corporations or institution type farms the person who makes the day-to-day decisions such as planting, harvesting, and marketing is considered the operator.

The list frame is a list of names of persons involved in farming operations and their corresponding addresses. In the list frame this address does not always correspond to the headquarters as in the area frame. The address in the list is the place where the person wants all correspondence to be mailed. Not all target population units are assigned to the list frame units, i.e. the list may be incomplete. Duplication may also occur within the list frame when one or more outside sources are combined to update and maintain the master list. Duplicated information can also be obtained for joint operations if each partner is on the list as one or a combination of names.

Due to this duplicated information on the list, rules must be developed to define and associate a target population unit with a sampling frame. Vogel [5] investigates problems in multiple frame applications using three different rules of association. Two of those rules will be used for the purpose of this paper to define the operator(s) of the population units.

The first rule we will use is the simpler of the two rules. The purpose of this rule is to eliminate any bias associated with determining if the operator of farm land is part of a joint operation. Rule 1 specifies that:

(a) The land operated singly by an individual name or in the name of a joint operation can only be associated with one frame unit.

(b) All joint operations in the target population will be handled using rules (i) thru (v) above.

The second set of rules primary purpose is to minimize the effect of partnership operations on the sampling errors. Rule 2 relies on some basic assumptions; viz.,

(a) Each partner in a partnership can report for the partnership operation whether contacted through the area or list sampling frames.

(b) Each partner can also report his individual operation if there is one.

(c) Each partner can correctly identify all of the other partners.

(d) Every partner that appears in the list frame will be identified.

The joint use of the two single frames is referred to as multiple frame sampling. Since the area frame is complete, every list sampling unit on the list frame can be mapped to a sampling unit in the area frame but not every name associated with the area frame is on the list frame. Thus multiple frame sampling presents us with the problem of determining the overlap between the sampling frames. It is necessary to determine which population units from the area frame could also have been obtained through the list frame. This determination is conducted by matching names. The problem is compounded when one or more names could be linked with the same operation.

UNBIASEDNESS

Now that all definitions, rules of association and assumptions have been stated, unbiased estimates for the population totals will be derived. These unbiased estimates naturally assume that all rules have been properly executed in the survey operations. We will consider estimators for the different mappings at the single frame level before combining frames. We assume simple random sampling.

The first frame to be considered is the area frame. The population units are farms as defined by Rule 1 and the sampling units are segments. As a means of identifying a farm with a unique segment we used the aforementioned headquarters rule. Let M = total number of population units (farms)

$$ {}_aX_i = \begin{cases} \text{Number of farm headquarters in the i-th sampling unit (segment)} \\ 0 \text{ if the i-th sampling unit (segment) contains no farms.} \end{cases} $$

Since the area frame is complete we have $M = \sum_{i=1}^{N_a} {}_aX_i$

where ${}_aN$ is the total number of sampling units (segments). To estimate M we use $\hat{M}$ where

$$\hat{M} = \sum_{i=1}^{{}_an} \frac{{}_aN}{{}_an} \, {}_aX_i$$

$$= \sum_{i=1}^{{}_aN} \frac{{}_aN}{{}_an} \, {}_aX_i \, \tau_i, \text{ where}$$

${}_an$ is the number of segments sampled and

$$\tau_i = \begin{cases} 1 \text{ if i-th frame unit is selected} \\ 0 \text{ otherwise} \end{cases}$$

and $E(\tau_i) = \frac{{}_an}{{}_aN}$

The expected value of $\hat{M}$ is clearly M.

The listed frame is a more complex problem than the area frame. In the list frame certain problems arise in mapping from the population units to the sampling units. The cases that will be considered are:

1. the list frame without joint operations or duplication under Rule 1.
2. the list frame with duplication under Rule 1.
3. the list frame with joint operations and duplication under Rule 2

Consider the list frame assuming it does not contain joint operations or duplication. The population unit is the farm and the sampling unit is a name and its corresponding address. Then there is a mapping from the target population units to the sampling frame units under Rule 1. Under Rule 1, which states that each name will report for itself, cases 1 and 2 can be shown to be unbiased.

Let ${}_LM$ = number of population units accounted for by the list frame.

Then

$${}_LM = \sum_{i=1}^{{}_LN} {}_LX_i,$$

where ${}_LX_i = \begin{cases} 1 \text{ if name = farm} \\ 0 \text{ if name} \neq \text{farm} \end{cases}$

and ${}_LN$ = total number of list frame units

Note: ${}_LM < M$ since the list is incomplete.

An estimate for ${}_LM$ based on a sample of size ${}_Ln$ is ${}_L\hat{M} = {}_LN \, {}_L\bar{X}_i$

$$= {}_LN \sum_{i=1}^{{}_LN} \frac{{}_LX_i}{{}_Ln}$$

$$= {}_LN \sum_{i=1}^{{}_LN} \frac{{}_LX_i}{{}_Ln} \tau_i$$

where τ_i is as defined before. Again ${}_L\hat{M}$ is an unbiased estimate of ${}_LM$.

The second case for which Rule 1 applies is the list frame with duplication. Rao [4] developed a procedure for handling duplication within the list where the number of times an operation can be selected is known. To develope the unbiasedness let

${}_LN$ = total number of list frame units

${}_LN'$ = total number of unique list frame units

${}_LA_i = \begin{cases} \text{number of times each unit is duplicated} \\ 1 \text{ if not duplicated, } i = 1, \ldots, {}_LN \end{cases}$

${}_LX_i = \begin{cases} 1 \text{ if farm = name} \\ 0 \text{ if farm} \neq \text{name, } i = 1, \ldots, {}_LN \end{cases}$

We then have

$${}_LM = \sum_{i=1}^{{}_LN} \frac{{}_LX_i}{{}_LA_i} = \sum_{i=1}^{{}_LN'} \frac{{}_LX_i}{{}_LA_i} \; {}_LA_i = \sum_{i=1}^{{}_LN} {}_LX_i$$

$$\therefore \; {}_L\hat{M} = \sum_{i=1}^{{}_Ln} \frac{{}_LN}{{}_Ln} \frac{{}_LX_i}{{}_LA_i} = \sum_{i=1}^{{}_LN} \frac{{}_LN}{{}_Ln} \frac{{}_LX_i}{{}_LA_i} \tau_i$$

where

$$\tau_i = \begin{cases} 1 \text{ if i-th frame unit is selected} \\ 0 \text{ if i-th frame unit is not selected,} \end{cases}$$

$i=1,\ldots, {}_LN$. Upon taking the expected value of ${}_L\hat{M}$ we obtain the desired unbiasedness of the estimator.

In contrast to Rule 1, Rule 2 allows for a population unit to be associated with more than one operator in the case of joint operations. Therefore, Rule 2 adds a new dimension to the problem of unbiased estimates by introducing another kind of duplication. As can be seen Rule 2 is more difficult to apply than Rule 1.

We will now consider the list frame with joint operations and duplications under Rule 2. Define ${}_LN, {}_LN'$ and ${}_LA_i$ as before and let

${}_LA'_i = \begin{cases} \text{number of times the population unit is} \\ \text{uniquely duplicated by different persons} \\ 1 \text{ otherwise, } \; i = 1, \ldots, {}_LN \end{cases}$

In determining ${}_LA'_i$ we are concerned with duplication of population units whereas ${}_LA_i$ was concerned with the duplication of frame units.

To see the use of ${}_LA_i$ and ${}_LA'_i$ consider the following situation for joint operations where the letters represent names with an address of persons who can report for some population unit:

Population	List Frame Units
1. A, B and C	1. B
2. D and E	2. A, B
	3. A
	4. C
	5. D
	6. A

The only duplication of frame units occurs between the third and sixth units. Therefore ${}_LA_i = 2$ for these two units and ${}_LA_i = 1$ for the remaining units in the frame.

${}_LA'_i$ is determined by the number of times a population unit is uniquely duplicated by different persons. The first population unit is duplicated by persons corresponding to the first, second, third, and fourth and sixth frame units. But the third and sixth units are duplicates so they will be counted only once. This then leaves 4 unique frame units associated with the first population unit. We then associate an ${}_LA'_i = 4$ for the five frame units associated with the first population unit. Since only one frame unit is associated with the second population unit it will have an ${}_LA'_i = 1$.

From the example above we can calculate ${}_LM$, where ${}_LX_i = 1$ for all frame units since each farm can be associated with a frame unit.

Then

$${}_LM = \sum_{i=1}^{6} \frac{{}_LX_i}{{}_LA_i\,{}_LA'_i} = 2$$

Therefore ${}_LM = 2$, which is the number of population units for this example. An estimate of ${}_LM$ is

$${}_L\hat{M} = \sum_{i=1}^{{}_Ln} \frac{{}_LN}{{}_Ln} \frac{{}_LX_i}{{}_LA_i\,{}_LA'_i}$$, which is clearly unbiased assuming a properly executed mapping has occurred.

Up to this point we have discussed two rules of association between the population units and the sampling units for single frames, i.e., the area frame and the list frame. We would now like to mention the multiple frame.

The procedures used rely on certain assumptions in application of the two previously stated rules (see Rule 1 and 2). Decision diagrams for both sets of rules are generally used to determine the nonoverlap between frames. Since the area frame is complete and the list frame is incomplete the expression for M, the total number of population units, is

$M = {}_{NOL}M + {}_{OL}M$, where

NOL = not on list and OL = on list. From this expression it is easily seen that ${}_{OL}M = {}_LM$, the total number of population units associated with the list.

$$M = {}_{NOL}M + {}_{OL}M = {}_{NOL}M = {}_LM$$
$$= {}_{NOL}M + P\,{}_{OL}M + Q_LM, \text{ where } P + Q = 1.$$

The expression ${}_{NOL}M + P_{OL}M$ is associated with the area frame and Q_LM is associated with the list frame.

SUMMARY

In this paper two rules for associating a target population unit with a sampling frame unit were presented. The first rule stated that each population unit reports for only its farm. The second rule stated that a frame unit can report for each population unit it is affiliated with. This is a drastic difference from the first rule in application. Much more work is needed in the form of checking the frame for the other members of joint operations and in the form of actually prorating the data. This additional work may lead to an increase in nonsampling errors for a survey.

The two rules were collectively applied to the area frame and the list frame under certain conditions. In both cases unbiased estimates of the total number of farms represented by each frame were obtained.

Additional rules should be developed and tested to associate the target population unit with the sampling frame unit. The importance of such considerations in surveys which required unbiased estimates is stressed.

REFERENCES

[1] Cochran, W.G., Discussion, International Institute of Statistics, 40 (1963), 538-539

[2] Deming, W.E., Discussion, International Institute of Statistics, 40 (1963), 543

[3] Hansen, M.H., Hurwitz, W.N., and Jabine, T.B., "THE USE OF IMPERFECT LISTS FOR PROBABILITY SAMPLING AT THE U.S. BUREAU OF CENSUS", International Institute of Statistics, 40 (1963), 497-517

[4] Rao, J.N.K., "SOME NONRESPONSE SAMPLING THEORY WHEN THE FRAME CONTAINS AN UNKNOWN AMOUNT OF DUPLICATION", Journal of the American Statistical Association, March 1968, 87-90

[5] Vogel, F.A., "SURVEYS WITH OVERLAPPING FRAMES - PROBLEMS IN APPLICATION", Proceedings of the Social Statistics Section, American Statistical Association, 1975

[6] "1976 June Enumerative and Multiple Frame Surveys - Interviewers' Manual' USDA

METHODS AND COSTS OF A 17-YEAR FOLLOW-UP STUDY*

Brian R. Clarridge, Linda S. Sheehy, Taissa S. Hauser, Robert M. Hauser
University of Wisconsin-Madison

The importance of finding ways to minimize attrition in longitudinal studies has been highlighted by two recent developments affecting social surveys. First, response rates in surveys of the general population have declined sharply in recent years, primarily because of increased numbers of refusals and persistent not-at-homes. Second, social science researchers have come to rely increasingly on common sources of survey data from large samples of representative populations. These suggest the importance of reducing those types of nonresponse which may be more tractable than refusal.

While reports of successful, large-scale and long-term panel surveys are not unknown, notable failures are also common. Major factors affecting tracing failure are the size, social integration, mobility, and dispersion of the sample and the elapsed time between waves of the survey. We think another factor, the intensity and ingenuity of the search procedure, is equally important, for halfhearted efforts to trace respondents perpetuate the persistent myth that uncontrollable factors prevent successful tracing operations. Too often the search is left to individuals who are not involved with or do not have a vested interest in the research and the quality of the data.

Over periods ranging from a year or two to ten or fifteen years the response rates in the better known surveys vary from close to 90 percent to as little as 20 percent. Although there have been several reports of successful tracing attempts, our description of the present effort may be useful in several ways: First, our tracing operation was carried out mainly during the calendar years of 1974-75 and our follow-up survey in the calendar year 1975. The recency of our experience strongly indicates that changes in receptivity to surveys have not made it impossible to carry out a successful large-scale follow-up panel study. Second, our sample differs in important ways from several of those for which successful follow-ups have been previously reported. It is very large (more than 10,000 persons) and geographically dispersed, and the elapsed time since the last direct or indirect contact with respondents ranged from ten to seventeen years. Third, we believe that finding people may be more a craft than a science. The description of successful searches may be the main means by which the craft may be improved, given the variation in the characteristics of samples and of information available about them from one panel survey to another. Fourth, we want to emphasize our impression that the organization and management of the search procedure plays an important part in its success--a part which is not adequately portrayed by the mere codification of information sources and search procedures.

We are reporting on a search for members of a sample of 10,317 men and women who were seniors in Wisconsin high schools in 1957. The post-high school experiences and achievements of this sample have been studied extensively by William H. Sewell and his associates of the University of Wisconsin. Data for these studies came from questionnaires filled out when students were seniors in high school, a 1964 postcard survey directed to the students' parents, Social Security earnings histories, Wisconsin tax records, and, most recently, a round of telephone interviews. Our tracing operation carried out in 1974 successfully located 97.4 percent of the original members of the sample. This figure includes 99 percent of the 9,007 persons for whom responses were obtained in the 1964 survey, and 86.2 percent of the 1,310 persons for whom no responses had been obtained in the 1964 survey. Ultimately, 88.6 percent of the members of the original sample were interviewed by telephone during 1975. This compares favorably to the 87.3 percent of the original 10,317 people for whom data were obtained in 1964. Those interviewed comprise 91.4 percent of the original sample who were not known to be deceased, disabled, or currently living outside the United States. The response rate is slightly higher for females than males--92.3 percent and 90.5 percent, respectively. The response rate is 93.5 percent among persons whose parents responded in 1964, and 77 percent among 1964 nonrespondents. This differential is mainly due to tracing failure. We believe that our success is in a large part attributable to our extensive use of the telephone, both during the search phase and during the interviewing process, as well as to certain features of the organization of the search procedure.

The full paper** includes a detailed exposition of our research design, together with a description of the organization, and a description of the procedures used in the tracing operation. It gives an accounting of the clerical time required and the percentage of cases found in various phases of the search. For a random tenth of the sample it reports differentials in tracing complexity and success, and in interview completion rates by sex, educational attainment, urban and rural residence, and in- and out-of-state residence.

We shall now review a few features of the

---

*This research was supported by a grant from the National Institute of Mental Health (M-06275), on which William H. Sewell and Robert M. Hauser are principal investigators. File management and computation were carried out using facilities of the Center for Demography and Ecology, the University of Wisconsin-Madison, which are supported by a Centers grant from the National Institute for Child Health and Human Development (HD-05876). The authors thank Professor William H. Sewell for his suggestions and support in revising this paper. We also thank Nancy Bode for her assistance with the computations.

**Working Paper 76-23, Center for Demography and Ecology, University of Wisconsin-Madison, Madison, Wisconsin, 53706.

search and its results. First, we examined the various types of identifying information which were available from the 1957 survey, the 1964 parent interview, or the various record data (Figure 1). We then began our trace with a telephone book search, first for the parent at his or her most recent address, and then for the respondent at his or her most recent address. These two steps led us directly or indirectly to close to 90 percent of the students. We would add that in many cases finding the respondent was not so easy as a single, direct look-up. For example, our telephone book search steps included coverage of listings in geographically contiguous areas and the calling of same last-name persons in the same locality.

Where the telephone book searches failed to produce successful leads we reviewed each case and tried one or more of the following leads: First, we tried occupational leads including names of employers, unions, licensing and professional organizations, or job titles which in some cases led us to places of employment. Second, we looked for alumni leads at colleges or universities which students attended, graduated from, or which they had at some time planned to attend. Third, we tried neighbors in those places where reverse-listed address directories existed. We looked for the telephone numbers of individuals living near the respondent at some previous time or near his parent at some previous time. Using this method we initially looked for persons at the same address, and then for persons next door or across the street, or in other nearby addresses. Fourth, we sometimes contacted high schools. Occasionally, they would keep registers of current addresses of their graduates; sometimes they would know the names of siblings; at other times we could find out from where transcripts had been requested; or occasionally a school secretary would have personal knowledge about the whereabouts of one of our intended respondents. Similarly, we were often able to obtain first-hand information about respondents by calling post offices--especially in small towns and rural areas. For men with military service we had some success using the military locator services which various branches of the armed forces maintain. Finally, as a last-ditch effort, we wrote a letter to those fellow classmates of our intended respondent who had already been interviewed. This took place more than a year after the bulk of our research, and it had a very low return rate, primarily because we had already found almost everyone in our sample. Where we did have a response to this letter, it was almost always successful in leading us to one of the persons whom we had otherwise failed to locate.

We closed only those cases in which we were able to ascertain the current name, address, and telephone number of the respondent from that persons' parent, spouse, or responsible child. Thus, we made a large number of confirming calls once we had learned the location from some other source. Ultimately, we know of only two possible occurrences of errors in identifying a respondent, and we think one of those was a clever refusal.

Since we found and interviewed such a high proportion of the sample, the response differentials were rather weak (Table 1). Men and persons of urban origin were harder to find and interview than women and persons of rural origin. Business leads were especially helpful in locating male college graduates, and spouses' names and employment characteristics in locating females. More than a third of the sample members were located with a single call, and two-thirds after no more than three calls, but the distribution was highly skewed to the right. Nine percent of the sample required 11 or more calls before we located them. Our considerable efforts to locate hard to find persons were well rewarded. Even in the hardest to find eight percent of the sample we ultimately interviewed 71 percent of the cases, and among those we attempted to call who were not dead, outside the U.S., or without a phone, the completion rate was more than 90 percent.

Finally, we note several organizational features of our research. First, we carried out a substantial pretest, just as we did before the production interviewing was undertaken. This helped us train personnel, develop search and record-keeping procedures, and anticipate problems with particular types of respondents. Second, the production tracing was carried out in highly stratified random tenths of the full sample. This smoothed the work flow, helped in training searchers, prevented a pile-up of hard-to-find cases, and enabled us to monitor and control costs without a complex accounting system. Third, our extensive use of the telephone, combined with the central review of cases by means of a continuous and complete call record, allowed us to maximize feedback about each case in real time and to adapt both our search strategy and tactics to this feedback. Finally, we treated the search procedure as inherently unreliable, and obtained a substantial return from verification of records, repeating of steps, and in some cases a complete recycling of failed cases.

It would be tempting to conclude from all this that our success in tracing the Wisconsin panel was an artifact of the restricted and favorable make-up of the sample. While that argument has some merit, we feel the lack of substantial response differentials across the major strata of our sample encourages the possibility of similar success under other conditions.

Table 1. Percentage Distribution of Final Status of Search by Sex, Education, Rural-Urban Residence in 1957, and Instate or Out-of-State Residence in 1964 and 1975*

	Inter-viewed	Re-fused	No Phone	De-ceased**	Outside USA	Not Found	Total	N
Male	89.22	4.59	1.83	1.15	1.15	2.06	100	(436)
Female	92.61	2.83	1.52	1.09	0.65	1.30	100	(460)
High School	91.40	3.95	2.09	1.16	0.23	1.16	100	(430)
Vocational or Some College	90.32	3.23	1.94	0.65	1.29	2.58	100	(310)
College Grad	91.03	3.85	0.0	1.92	1.92	1.28	100	(156)
1957 Rural***	95.22	2.39	1.19	0.30	0.30	0.60	100	(335)
1957 Urban	88.41	4.46	1.96	1.60	1.25	2.32	100	(561)
1964 Instate	90.98	4.37	1.60	1.16	0.15	1.75	100	(687)
1964 Out-of-State	90.91	1.44	1.91	0.96	3.35	1.44	100	(209)
1975 Instate	92.40	4.50	1.55	1.55**	0.0	0.0	100	(645)
1975 Out-of-State	91.63	1.67	2.09	0.0	3.35	1.26	100	(239)
Not Found	0.0	0.0	0.0	0.0	0.0	100	100	(12)

*This table represents the final outcome of the search,thus there are 3 individuals who were found, then lost again,bringing the total in the Not Found column to 15 rather than the 12 indicated in Tables 2, 3, and 4.

**In all tables the deceased were coded INSTATE in 1975.

***Cities with a population of less than 2500 were coded as rural.

Figure 1. Identifying information for 1964 respondents and nonrespondents

Identifying Information	1964 Respondents	1964 Nonrespondents
Parent's and student's 1957 address (the same household)	X	X
Parent's 1964 address	X	-
Student's 1964 address	X	-
Schools the student attended or graduated from prior to 1964	X	-
Student's 1964 occupation	X	-
Student's marital status in 1964	X	-
Female student's spouse's occupation	X	-
Student's status with respect to military service	X	-
Background Information		
Parents' education	X	X
High school rank	X	X
Student's occupational aspirations	X	X
Parent's 1957 occupation	X	X

WARNER'S RANDOMIZED RESPONSE TECHNIQUE: THE TWO SENSITIVE QUESTION CASE

Robert P. Clickner and Boris Iglewicz, Temple University

1. Introduction

Surveys, particularly those involving sensitive questions, have always been plagued by biases caused by non-response and untruthful responses. The first randomized response scheme designed to reduce these biases was developed by Warner [2]. Since then a number of generalizations and variations of Warner's technique have been developed. The recent survey paper by Horvitz, Greenberg and Abernathy [1] summarizes much of this work. The previous randomized response papers discuss the estimation of parameters and efficiencies of the randomized response techniques as compared to direct question surveys, but only in the context of analysis based on one sensitive question at a time. In actual surveys one is not only interested in the analysis of single characteristics, but also in joint estimates of a number of characteristics. Actually one is interested in obtaining as much useful information as possible and the popularity of 2×2 contingency tables reflects the need for cross-tabulation analysis of survey data.

In this paper we consider the extension of Warner's scheme to two sensitive questions. The maximum likelihood estimators of the joint and marginal probabilities are derived and their means, variances and covariances are obtained. Some properties of these estimators are explored and recommendations concerning choice of randomization parameters are made. Also considered are the efficiencies of these estimators relative to direct questioning; assuming both zero and positive probabilities of untruthful responses.

2. Statement of the Problem

The problem is to simultaneously estimate the proportion of the population who possess either or both of two sensitive characteristics using Warner's [2] randomized response scheme. If A and B are the sensitive characteristics let

π_A = P(a person is an A),

π_B = P(a person is a B),

π_{AB} = P(a person is both an A and a B).

We need only estimate π_A, π_B and π_{AB}. Once these three probabilities are estimated, all the other joint, marginal, and conditional probabilities concerning A and B can be easily estimated.

Warner's scheme, extended to the two question case, proceeds as follows. The respondent is given a random device (e.g., a spinner, deck of cards, box of colored marbles, etc.) with which to choose one of the two statements

1-a. I am an A,
1-b. I am not an A.

The device selects 1-a with probability p_1 and 1-b with probability $\bar{p}_1 = 1-p_1$ (Given any probability θ, we will always use the notation $\bar{\theta}$ for $1-\theta$). Without revealing to the interviewer which statement has been chosen, the respondent answers "yes" or "no" according to the statement selected and to his actual status with respect to the characteristic A. This procedure is then repeated with the statements

2-a. I am a B,
2-b. I am not a B,

and with a second random device which selects 2-a with probability p_2 and 2-b with probability $\bar{p}_2$. We need not have $p_1 = p_2$. Hence the information received from each respondent is one of the four pairs: yes-yes, yes-no, no-yes, no-no. We will code the responses as 1 = "yes" and 0 = "no."

Let λ_{ij} = P(response i on the first question and response j on the second question), $i = 0, 1$; $j = 0, 1$. Then writing $\bar{A}$ for "not A" and $\bar{B}$ for "not B," we obtain

$$\begin{aligned}\lambda_{11} &= p_1p_2P(AB)+p_1\bar{p}_2P(A\bar{B})+\bar{p}_1p_2P(\bar{A}B)+\bar{p}_1\bar{p}_2P(\bar{A}\bar{B}) \\ &= p_1p_2\pi_{AB}+p_1\bar{p}_2(\pi_A-\pi_{AB})+\bar{p}_1p_2(\pi_B-\pi_{AB}) \\ &\quad + \bar{p}_1\bar{p}_2(1-\pi_A-\pi_B-\pi_{AB}) \\ &= \pi_A\bar{p}_2(p_1-\bar{p}_1)+\pi_B\bar{p}_1(p_2-\bar{p}_2)+\pi_{AB}(p_1-\bar{p}_1)(p_2-\bar{p}_2) \\ &\quad +\bar{p}_1\bar{p}_2, \qquad (2.1)\end{aligned}$$

$$\begin{aligned}\lambda_{10} &= \pi_Ap_2(p_1-\bar{p}_1)-\pi_B\bar{p}_1(p_2-\bar{p}_2)-\pi_{AB}(p_1-\bar{p}_1)(p_2-\bar{p}_2) \\ &\quad +\bar{p}_1p_2, \qquad (2.2)\end{aligned}$$

$$\begin{aligned}\lambda_{01} &= -\pi_A\bar{p}_2(p_1-\bar{p}_1)+\pi_Bp_1(p_2-\bar{p}_2)-\pi_{AB}(p_1-\bar{p}_1)(p_2-\bar{p}_2) \\ &\quad +p_1\bar{p}_2 \qquad (2.3)\end{aligned}$$

$$\begin{aligned}\lambda_{00} &= -\pi_Ap_2(p_1-\bar{p}_1)-\pi_Bp_1(p_2-\bar{p}_2)+\pi_{AB}(p_1-\bar{p}_1)(p_2-\bar{p}_2) \\ &\quad +p_1p_2. \qquad (2.4)\end{aligned}$$

3. Estimation of π_A, π_B, π_{AB}; Test for Independence

Consider a sample of n respondents and let X_{ij}, $i = 0, 1$; $j = 0, 1$, be the number of respondents who respond i to question 1 and j to question 2. The joint distribution of X_{11}, X_{10}, X_{01} and X_{00} is multinomial (n; λ_{11}, λ_{10}, λ_{01}, λ_{00}). The maximum likelihood estimator (MLE) of λ_{ij} is $\hat{\lambda}_{ij} = X_{ij}/n$. Using equations (2.1)-(2.4) and some algebra one obtains the MLE's of π_A, π_B, π_{AB} to be

$$\hat{\pi}_A = (\hat{\lambda}_{11}+\hat{\lambda}_{10}-\bar{p}_1)/(p_1-\bar{p}_1),$$

$$\hat{\pi}_B = (\hat{\lambda}_{11}+\hat{\lambda}_{01}-\bar{p}_2)/(p_2-\bar{p}_2),$$

$$\hat{\pi}_{AB} = [\hat{\lambda}_{11}(p_1p_2-\bar{p}_1\bar{p}_2)-\hat{\lambda}_{10}\bar{p}_2-\hat{\lambda}_{01}\bar{p}_1+\bar{p}_1\bar{p}_2] / (p_1-\bar{p}_1)(p_2-\bar{p}_2),$$

respectively. Note that $\hat{\pi}_A$ and $\hat{\pi}_B$ are exactly the same estimates as found in Warner [2].

The maximum likelihhod estimates of π_A, π_B, π_{AB} are unbiased and have the following variances and covariances:

$$V(\hat{\pi}_A) = [\pi_A\bar{\pi}_A+f(p_1)]/n,$$

$$V(\hat{\pi}_B) = [\pi_B\bar{\pi}_B+f(p_2)]/n,$$

$$V(\hat{\pi}_{AB}) = [\pi_{AB}\bar{\pi}_{AB}+\pi_A f(p_2)+\pi_B f(p_1) +f(p_1)f(p_2)]/n,$$

$$Cov(\hat{\pi}_A,\hat{\pi}_B) = (\pi_{AB}-\pi_A\pi_B)/n,$$

$$Cov(\hat{\pi}_A,\hat{\pi}_{AB}) = [\pi_{AB}\bar{\pi}_A+\pi_B f(p_1)]/n,$$

$$Cov(\hat{\pi}_B,\hat{\pi}_{AB}) = [\pi_{AB}\bar{\pi}_B+\pi_A f(p_2)]/n,$$

where $f(p) = p\bar{p}/(p-\bar{p})^2$.

Let $\lambda_{i\cdot} = \lambda_{i1}+\lambda_{i0}$, $i = 0,1$ and $\lambda_{\cdot j} = \lambda_{1j}+\lambda_{0j}$, $j = 0,1$. Then we have $\lambda_{11} = \lambda_{1\cdot}\lambda_{\cdot 1}$ if and only if $\pi_{AB} = \pi_A\pi_B$. That is, the responses to questions 1 and 2 are independent if and only if the characteristics A and B are independent. It follows that we can test the independence of A and B by applying the ordinary χ^2 test to the randomized responses. Specifically we can use the test statistic

$$X^2 = \sum_{i,j}(X_{ij}-n\hat{\lambda}_{i\cdot}\hat{\lambda}_{\cdot j})^2/n\hat{\lambda}_{i\cdot}\hat{\lambda}_{\cdot j},$$

where $\hat{\lambda}_{i\cdot} = \hat{\lambda}_{i1}+\hat{\lambda}_{i0}$ and $\hat{\lambda}_{\cdot j} = \hat{\lambda}_{1j}+\hat{\lambda}_{0j}$.

4. Estimation of π_{AB} Only

Occasions may arise in which one is only interested in estimating π_{AB}. In such situations the question arises as to which is the better procedure--the two-question procedure described in Section 2 or Warner's original procedure applied to A∩B, that is, asking the respondent to answer one of the questions

3-a. I am both an A and a B,
3-b. I am a not-A or a not-B.

If question 3-a is selected with probability p and 3-b with probability $\bar{p}$, the resulting estimator $\tilde{\pi}_{AB}$ (see Warner [2]) of π_{AB} is unbiased and has variance $V(\tilde{\pi}_{AB}) = [\pi_{AB}\bar{\pi}_{AB}+f(p)]/n$. It is interesting to note that neither procedure is uniformly better than the other. In fact, if we let $p_1 = p_2 = p$, (for simplicity, and in accordance with Section 5), then

$$V(\hat{\pi}_{AB})/V(\tilde{\pi}_{AB}) = 1+f(p)[\pi_A+\pi_B+f(p)-1]/V(\tilde{\pi}_{AB}),$$

which is less than one when $\pi_A+\pi_B+f(p) < 1$, and this occurs when π_A, π_B and $1/2-|1/2-p|$ are all relatively close to zero. Both procedures provide relatively little confidentiality under these circumstances (p is close to zero or one), but the two question approach is revealing for all the responses in A∪B, while the one question case is revealing only for those in A∩B. Thus, the smaller variance of $\hat{\pi}_{AB}$ is likely to be somewhat offset by a greater likelihood of nonresponse and/or untruthful responses.

Some comparisons between $V(\hat{\pi}_{AB})$ and $V(\tilde{\pi}_{AB})$ are given in Table 1. Note that either variance can be considerably larger than the other, but $V(\tilde{\pi}_{AB})$ is the larger only when sampling for rare characteristics with a small p. More typically, $V(\hat{\pi}_{AB})$ is considerably larger than $V(\tilde{\pi}_{AB})$. This illustrates that as one goes from a single sensitive question to a two sensitive question analysis, there can be a great increase in the variance of the estimate of the joint probability. Such a comparison will show even greater increases in variances when more than two sensitive questions are asked. This can be easily seen for the special case of independent characteristics.

Table 1. $V(\hat{\pi}_{AB})$ and $V(\tilde{\pi}_{AB})$ for Selected Values of π_A, π_B, π_{AB}, and p

π_A	π_B	π_{AB}	p	$nV(\hat{\pi}_{AB})$	$nV(\tilde{\pi}_{AB})$
.01	.0075	.0025	.4	36.107	6.002
.01	.0075	.0025	.1	0.025	0.143
.04	.0300	.0100	.4	36.430	6.010
.04	.0300	.0100	.1	0.040	0.151
.16	.0400	.0133	.4	37.213	6.013
.16	.0400	.0133	.1	0.061	0.154
.64	.3200	.1067	.4	41.855	6.095

5. Efficiency of the Estimators; Choice of p_i's

For the two sensitive question case one is usually interested in estimating π_A, π_B and π_{AB} individually. Hence a reasonable measure of the efficiency of the estimation procedure is the trace of the variance-covariance matrix; that is, the quantity

$$v(p_1,p_2) = n[V(\hat{\pi}_A)+V(\hat{\pi}_B)+V(\hat{\pi}_{AB})]. \tag{5.1}$$

It is to be noted that $v(p_1,p_2)$ depends on p_1 and p_2 only through the function $f(p)$. Further, $f(p)$ has the properties: $f(0) = 0$; $f(\bar{p}) = f(p)$, $0 \le p \le 1$; $df/dp > 0$, $0<p<1/2$; and $f(p) \to \infty$ as $p \to 1/2$. It follows that $\min_{p_1,p_2} v(p_1,p_2) = v(0,0)$, which is the measure obtained for the direct question approach.

Since maximum statistical efficiency cannot be achieved without destroying all confidentiality, one could take the approach of selecting p_1 and p_2 to achieve a given preassigned efficiency. That is, given efficiency $1/r$, $r>1$, select p_1 and p_2 to satisfy

$$v(p_1,p_2) = rv(0,0). \tag{5.2}$$

This is one equation in two unknowns and therefore has infinitely many solutions. Without loss of generality assume $0<p_1<1/2$ and $0<p_2<1/2$. Then, solving (5.2) for p_2 as a function of p_1 we obtain

$$p_2 = [1-(4f_2+1)^{-1/2}]/2,$$

where

$$f_2 = [(r-1)v(0,0)-f_1(1+\pi_B)]/[f_1+1+\pi_A],$$

and $f_1 = f(p_1)$. Routine application of the calculus establishes that p_2 is a continuous, strictly decreasing and concave function of p_1.

An obvious and convenient solution for equation (5.2) can be obtained by choosing $p_1 = p_2$ (or $p_1 = \overline{p}_2$, because of the symmetry of $f(p)$). This solution maximizes $\min(p_1,p_2)$. That is, it gives the greatest protection to the respondent on each question for a given efficiency. It was conjectured that there may occasionally be an advantage in choosing $p_1 \neq p_2$, especially when π_A differed greatly from π_B. It was felt that a small decrease in one of the p's would lead to a far larger increase in the other. This did not happen as can be seen from the entries of Table 2, in which r = 10. For example, when $\pi_A = .64$ and $\pi_B = .01$, with $\pi_{AB} = .00125$, then $p_1 = p_2 = .237$ is one solution to (5.2). Another solution is $p_1 = .220$ and $p_2 = .249$. That is, a sacrifice in p of .017 on question 1 yields a gain of .012 on question 2. There is little advantage in terms of protecting the respondent's privacy, in choosing the latter solution over $p_1 = p_2 = .237$. In sum, the solution $p_1 = p_2$ is quite feasonable and little can be gained by choosing $p_1 \neq p_2$.

Let us now consider the solution $p_1 = p_2 = p$ (say). Table 3 gives solutions of $v(p,p) = rv(0,0)$ for selected values of 1/r, and π_A, π_B and π_{AB}. Typical choices for p recommended in the literature (see [2] for details) are in the range from .2 to .3. In the two sensitive question case, Table 3 shows that this will typically result in a loss in efficiency of at least 70%, compared to direct questioning and assuming all responses are truthful.

Table 2. Values of p_1 and p_2 that Achieve 10% Efficiency for Selected π_A, π_B and π_{AB}

π_A	.16		.32		.64	
π_B	.16		.08		.01	
π_{AB}	.04		.04		.00125	
	p_1	p_2	p_1	p_2	p_1	p_2
	.000	.346	.000	.342	.000	.301
	.036	.342	.036	.339	.034	.297
	.069	.338	.071	.335	.068	.293
	.104	.333	.107	.329	.101	.288
	.138	.325	.142	.323	.135	.281
	.173	.316	.178	.313	.169	.271
	.208	.302	.213	.300	.203	.258
	.242	.282	.249	.281	.220	.249
	.263	.263	.267	.267	.237	.237
	.277	.249	.284	.249	.271	.206
	.311	.185	.320	.186	.304	.148
	.346	.000	.356	.000	.339	.000

Table 3. Values of p Required for Given Efficiencies for Selected π_A, π_B and π_{AB}

			Efficiency (1/r)			
π_A	π_B	π_{AB}	.8	.4	.2	.1
.05	.05	.0125	.012	.061	.122	.187
.10	.05	.0250	.018	.082	.153	.219
.20	.15	.0750	.037	.131	.211	.273
.25	.05	.0375	.027	.112	.190	.255
.25	.25	.0625	.038	.142	.223	.284
.25	.25	.2500	.047	.163	.244	.301
.40	.05	.0250	.029	.118	.197	.262
.55	.25	.1250	.042	.152	.234	.294
.75	.05	.0250	.022	.096	.172	.240
.75	.70	.5250	.041	.150	.234	.295

6. Effects of Untruthful Responses

Randomized response schemes are designed to reduce bias due to nonresponse or lying to protect one's privacy. To investigate the effects of such lying let

t_A = P(an A tells the truth about being an A),

and define t_B and t_{AB} similarly. Assume further, that persons not in a sensitive group will not claim to be members of such a group. Then for either randomized response or the direct question approach we have

$$\nu_A = E(\hat{\pi}_A) = \pi_A t_A \quad (6.1)$$

$$\nu_B = E(\hat{\pi}_B) = \pi_B t_B \quad (6.2)$$

$$\nu_{AB} = E(\hat{\pi}_{AB}) = \pi_{AB} t_{AB} \quad (6.3)$$

and the biases are therefore:

$$b(\hat{\pi}_A) = \pi_A \overline{t}_A \quad (6.4)$$

$$b(\hat{\pi}_B) = \pi_B \overline{t}_B \quad (6.5)$$

$$b(\hat{\pi}_{AB}) = \pi_{AB} \overline{t}_{AB} \quad (6.6)$$

It should be noted that the values of the t's in equations (6.1)-(6.6) will usually differ for the direct question case as compared to randomized response. Hopefully, the t's will have higher values for the randomized response scheme.

We will compare our randomized response scheme with direct questioning by using the sum of the mean squared errors (MSE):

$$M_t(p_1,p_2) = v(p_1,p_2)+b^2(\hat{\pi}_A)+b^2(\hat{\pi}_B)+b^2(\hat{\pi}_{AB})$$

where $M_t(0,0)$ is the measure obtained by direct questioning and $v(p_1,p_2)$ is given by equation (5.1).

Tables 4-7 give values of the sum of the biases ("bias") and $M_t(p_1,p_2)$ for selected values of t and $p_1 = p_2 = p$. Notice that for truthful responses by both methods, the randomized response technique leads to far larger values of MSE then does direct questioning. Consider now the case $\pi_A = .16$, $\pi_B = .12$, $\pi_{AB} = .04$, $p = .3$, $n = 1000$, and truthful responses for the Warner scheme while $t_A = .7$, $t_B = .6$, $t_{AB} = .5$ for direct questioning. Then $M_t(0,0) = .0051939$ while $M_t(.3,.3) = .00499$. So in this extreme case the Warner scheme is superior to direct questioning. If we consider the above example with $t_A = .9$, $t_B = .7$, $t_{AB} = .7$ then $M_t(0,0) = .0019234$ and direct questioning would be superior. In summary, the Warner scheme is superior to direct questioning only when randomized responses would produce considerable increases in the rate of truthful responses. It should be noted that in the above discussion we have excluded consideration of different rates of nonresponse resulting from using the two approaches.

Table 4. Effects of Lying on MSE; $n = 100$, $\pi_A = .04$, $\pi_B = .01$, $\pi_{AB} = .00667$

t_A	t_B	t_{AB}	bias	$M_t(0,0)\times10^6$	$M_t(p,p)\times10^6$	
					p = .3	p = .1
1.0	1.0	1.0	0.0	549	44682	3630
1.0	0.9	0.8	.00233	529	44648	3608
0.9	0.7	0.7	.00900	492	44532	3563
0.7	0.6	0.5	.01933	536	44458	3594
0.6	0.4	0.2	.02733	608	44451	3658

Table 5. Effects of Lying on MSE; $n = 1000$, $\pi_A = .04$, $\pi_B = .01$, $\pi_{AB} = .00667$

t_A	t_B	t_{AB}	bias	$M_t(0,0)\times10^7$	$M_t(p,p)\times10^7$	
					p = .3	p = .1
1.0	1.0	1.0	0.0	549	44682	3630
1.0	0.9	0.8	.00233	554	44674	3633
0.9	0.7	0.7	.00900	753	44794	3824
0.7	0.6	0.5	.01933	2076	45999	5134
0.6	0.4	0.2	.02733	3492	47336	6541

Table 6. Effects of Lying on MSE; $n = 100$, $\pi_A = .16$, $\pi_B = .12$, $\pi_{AB} = .04$

t_A	t_B	t_{AB}	bias	$M_t(0,0)\times10^6$	$M_t(p,p)\times10^6$	
					p = .3	p = .1
1.0	1.0	1.0	0.0	2784	49935	6188
1.0	0.9	0.8	.020	2825	49819	6212
0.9	0.7	0.7	.064	3970	50439	7301
0.7	0.6	0.5	.116	6867	52758	10136
0.6	0.4	0.2	.168	11708	57075	14921

Table 7. Effects of Lying on MSE; $n = 1000$, $\pi_A = .16$, $\pi_B = .12$, $\pi_{AB} = .04$

t_A	t_B	t_{AB}	bias	$M_t(0,0)\times10^7$	$M_t(p,p)\times10^7$	
					p = .3	p = .1
1.0	1.0	1.0	0.0	2784	49935	6188
1.0	0.9	0.8	.020	4697	51691	8084
0.9	0.7	0.7	.064	19234	65703	22565
0.7	0.6	0.5	.116	51939	97830	55208
0.6	0.4	0.2	.168	104444	149811	107657

References

[1] Horvitz, D. G., Bernard G., and Abernathy, J. R., "Recent Developments in Randomized Response Design," in J. N. Srivastava, ed., A Survey of Statistical Design and Linear Models, New York: North-Holland Publishing Company, 1975, 271-85.

[2] Warner, S. L., "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (March 1965), 63-9.

A CONTINGENCY TABLE MODELING APPROACH TO USING PARITY PROGRESSION RATIOS IN THE ANALYSIS OF THE NUMBER OF CHILDREN EVER BORN

Lester R. Curtin, I. Joel Frockt, and Gary G. Koch
University of North Carolina at Chapel Hill

1. Introduction

Recently, Namboodiri (1972,1974), Simon (1975a,1975b), and others have empirically studied fertility as a sequential decision-making process. It has been observed that a number of factors may have different effects at different birth orders. In particular, income may have both positive and negative effects, and the net effect of income on mean fertility or total fertility would then depend on the parity composition of the sample. For this reason, analysis of the effects of several factors may be obscured if the dependent variable is chosen to be some aggregate variable such as the mean number of children ever born (CEB).

Data on the number of CEB can be used to study the sequential thesis of fertility. Analytical methods which can be employed include regression analysis and discriminant function analysis. A contingency table modeling approach can also be used on the same type of data. This paper indicates how weighted least squares fitting procedures can be used to analyze parity progression ratios derived from the number of CEB. An example looks at the effects of religion, education, and income on the parity progression ratios.

2. Data

The data for the example are from the 1965 National Fertility Study (Princeton). The twelve sub-populations of white women are based on religion (Catholic, non-Catholic), education (less than high school, high school or better), and husband's income (low, medium, high). All women were in their first marriage and have had fifteen years or more exposure to the risk of pregnancy since marriage. As such, these women have effectively completed their childbearing experience.

3. Analysis

Data on the number of CEB for cohorts of women who have reached the end of their childbearing experience can be formulated in terms of a contingency table. Let n_{ij} be the number of women in the i-th sub-population (i=1,2,...,s) with parity j (j=0,1,...,r). The proportion of women in the i-th sub-population with parity j is

$p_{ij} = n_{ij}/n_i$, where $n_i = \sum_{j=0}^{r} n_{ij}$, and an unbiased estimate of the mean number of CEB is

$\hat{\mu}_i = \sum_{j=1}^{r} j\, p_{ij}$. An algebraically equivalent estimate can be derived from the parity progression ratios.

The parity progression ratios are defined for women in the i-th sub-population as

$a_{i0} = p_{i1}+p_{i2}+\ldots+p_{ir}$,

$a_{i1} = (p_{i2}+p_{i3}+\ldots+p_{ir})/(p_{i1}+p_{i2}+\ldots+p_{ir})$, ...,

$a_{ik} = (p_{ik+1}+\ldots+p_{ir})/(p_{ik}+p_{ik+1}+\ldots+p_{ir})$, ...,

$a_{ir-1} = (p_{ir})/(p_{ir-1}+p_{ir})$. Under the assumption that fertility has been constant, the k-th parity progression ratio for the i-th sub-population, a_{ik}(k=0,1,...,r-1), can be interpreted as the conditional probability of having a (k+1)-th birth. It can be seen that $a_{i0}a_{i1} = p_{i2}+\ldots+p_{ir}$, until $a_{i0}a_{i1},\ldots a_{ir-1} = p_{ir}$. Thus

$a_{i0}+a_{i0}a_{i1}+\ldots+a_{i0}a_{i1}\ldots a_{ir-1} = \sum_j j p_{ij} = \hat{\mu}_i$,

the estimate of the mean number of CEB. A closed form expression for the variance of a_{ik} is not readily available, but it has been asserted that $\mathrm{cov}(a_{ik},a_{ik'}) = 0$, $k \neq k'$.

The compound function formulation of Forthofer and Koch (1974) can be used to calculate the parity progression ratios and to obtain estimates of their covariance structure. The GSK weighted least squares method outlined in Grizzle, Starmer, Koch (1969) can now be applied to several models for the parity progression ratios. One strategy is to fit an incremental model to the parity classes within each of the sub-populations. A second strategy is to fit a factorial model across sub-populations for each parity class. The two models can be combined to account for both the variation across parities and the variation across sub-populations. For this method, an incremental model is fit across parities to the differences within the effects in the factorial model. For the data in Table 1, with the twelve sub-populations, this model has the parameterization displayed in Table 2.

The model is then reduced. The estimated parameters for the reduced model are reported in Table 4. The goodness of fit for the reduced model is summarized in Table 5. A more complete documentation of the analysis is given in Curtin et al (1976).

4. Discussion

The data used in the example are somewhat limited in scope but some interesting results can be noted by examining the predicted values in Table 6 which were obtained via the weighted least squares procedures. In general, the probability of an additional birth is greater for Catholics than for non-Catholics. However, for high parities, the low income, low education, non-Catholics have a greater probability of an additional birth than the corresponding sub-population of Catholics.

Education has no effect for the 0→1 parity transition and a positive effect for the 1→2 transition. Here a positive effect means that increased education is related to an increased probability of an additional birth. For the transition 2→3, and for higher parity transitions, education has a negative effect for all sub-populations except for the Catholic-high income sub-population which has a positive education effect for the transitions 1→2, 2→3, and 3→4 and a negative effect for the transitions 4→5 and 5→6. The education effect becomes more negative (less positive) as parity increases.

At all parities and sub-populations, income has a negative effect; that is, additional income decreases the probability of an additional birth, with the exception of low education Catholics at high parities. For this group, the probability of an additional birth decreases as income goes from low to medium, but then increases greatly for the high income sub-population at parities four or five. The income effects for non-Catholics increase as parity increases. No such generalization can be made for Catholics as there is substantial fluctuation across parities for the income effects.

ACKNOWLEDGMENTS

This research was in part supported by the U. S. Bureau of the Census through Joint Statistical Agreement JSA 76-93. The authors would like to thank Bea Parker for her conscientious typing of this manuscript.

REFERENCES

Bernhardt, Eva M. (1972). Fertility and economic status--Some recent findings on differentials in Sweden. *Population Studies 26*, 175-184.

Curtin, L.R., Frockt, I.J., and Koch, G.G. (1976). On the mixed effects of income and education on fertility. Submitted to *Demography*.

Grizzle, J.E., Starmer, C.F., Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrics 25*, 489-504.

Forthofer, R.N. and Koch, G.G. (1973). An analysis for compound functions of categorical data. *Biometrics 29*, 143-157.

Namboodiri, K. (1972). Some observations on the economic framework for fertility analysis. *Population Studies 26*, 185-206.

Namboodiri, K. (1974). Which couples at given parities expect to have additional births. *Demography 11*, 45-56.

Seiver, D.A. (1976). Which couples at given parities expect to have additional births. Paper presented at the annual meeting of the Population Association of America, Montreal, Canada.

Simon, Julian (1975a). The mixed effects of income upon successive births may explain the convergence phenomenon. *Population Studies 29*, 109-122.

Simon, Julian (1975b). Puzzles and further explorations in the interrelationships of successive births with husband's income, spouses' education, and race. *Demography 12*, 259-274.

TABLE 1

PARITY DISTRIBUTION BY RELIGION, EDUCATION, AND HUSBAND'S INCOME

Sub-population			Parity 0	1	2	3	4	5	6 or more	Total
Cath	less	low	3	6	19	11	17	10	18	84
Cath	less	med	5	15	12	18	21	8	12	91
Cath	less	high	2	2	7	4	2	1	7	25
Cath	more	low	2	2	6	6	9	3	6	34
Cath	more	med	4	2	34	22	15	15	13	105
Cath	more	high	7	18	30	17	20	9	13	114
Non	less	low	23	36	63	49	49	29	161	410
Non	less	med	15	34	63	52	28	24	35	251
Non	less	high	4	11	19	20	7	10	7	78
Non	more	low	14	12	34	31	17	11	23	142
Non	more	med	20	34	100	69	48	19	12	302
Non	more	high	33	47	108	99	50	20	12	369

TABLE 2

PARAMETERIZATION OF THE INCREMENTAL-FACTORIAL MODEL
WITHIN EACH PARITY CLASS

Source of Variation	Estimated Incremental Parameter	Indicator Variable
Mean	b_1	$x_1 = 1$ always
Main effect: Religion (R)	b_2	$x_2 = \begin{cases} 1 & \text{Catholic} \\ -1 & \text{Non-Catholic} \end{cases}$
Main effect: Education (E)	b_3	$x_3 = \begin{cases} 1 & \text{less than high school} \\ -1 & \text{high school of better} \end{cases}$
Main effect: Income (I)	b_4, b_5	$x_4 = \begin{cases} 1 & \text{low} \\ 0 & \text{medium} \\ -1 & \text{high} \end{cases}$, $x_5 = \begin{cases} 0 & \text{low} \\ 1 & \text{medium} \\ -1 & \text{high} \end{cases}$
Interaction: R × E	b_6	$x_6 = x_2 x_3$
Interaction: R × I	b_7, b_8	$x_7 = x_2 x_4$, $x_8 = x_2 x_5$
Interaction: E × I	b_9, b_{10}	$x_9 = x_3 x_4$, $x_{10} = x_3 x_5$
Interaction: R × E × I	b_{11}, b_{12}	$x_{11} = x_2 x_3 x_4$, $x_{12} = x_2 x_3 x_5$

TABLE 3

OBSERVED PARITY PROGRESSION RATIOS AND MEAN NUMBER CEB
WITH THEIR STANDARD ERRORS

Sub-population			Parity 0	1	2	3	4	5	Not Truncated Mean	Truncated Mean
Cath	less	low	0.964 (0.020)	0.926 (0.029)	0.747 (0.050)	0.804 (0.053)	0.622 (0.072)	0.643 (0.091)	3.929 (0.252)	3.607 (0.192)
Cath	less	med	0.945 (0.024)	0.826 (0.041)	0.831 (0.044)	0.695 (0.060)	0.488 (0.078)	0.600 (0.110)	3.264 (0.204)	3.176 (0.183)
Cath	less	high	0.920 (0.054)	0.913 (0.059)	0.667 (0.103)	0.714 (0.121)	0.800 (0.126)	0.875 (0.117)	3.840 (0.599)	3.320 (0.407)
Cath	more	low	0.942 (0.040)	0.938 (0.043)	0.800 (0.073)	0.750 (0.088)	0.500 (0.118)	0.667 (0.157)	3.618 (0.330)	3.500 (0.296)
Cath	more	med	0.962 (0.019)	0.980 (0.014)	0.656 (0.048)	0.661 (0.059)	0.652 (0.073)	0.464 (0.094)	3.375 (0.236)	3.324 (0.122)
Cath	more	high	0.939 (0.022)	0.832 (0.036)	0.663 (0.050)	0.712 (0.059)	0.524 (0.077)	0.591 (0.105)	3.051 (0.182)	2.912 (0.162)
Non	less	low	0.944 (0.011)	0.907 (0.015)	0.820 (0.020)	0.830 (0.022)	0.795 (0.026)	0.847 (0.026)	4.798 (0.161)	3.941 (0.100)
Non	less	med	0.940 (0.015)	0.856 (0.023)	0.688 (0.032)	0.626 (0.041)	0.678 (0.050)	0.593 (0.064)	3.457 (0.118)	3.020 (0.111)
Non	less	high	0.949 (0.025)	0.851 (0.041)	0.698 (0.052)	0.545 (0.075)	0.708 (0.093)	0.412 (0.119)	3.079 (0.244)	2.936 (0.186)
Non	more	low	0.901 (0.025)	0.906 (0.026)	0.707 (0.042)	0.622 (0.054)	0.667 (0.066)	0.676 (0.080)	3.190 (0.178)	3.056 (0.154)
Non	more	med	0.934 (0.014)	0.879 (0.019)	0.597 (0.031)	0.534 (0.041)	0.390 (0.055)	0.387 (0.087)	2.679 (0.087)	2.649 (0.082)
Non	more	high	0.911 (0.015)	0.860 (0.019)	0.626 (0.028)	0.453 (0.037)	0.390 (0.054)	0.375 (0.086)	2.542 (0.077)	2.526 (0.074)

TABLE 4

ESTIMATED INCREMENTAL-FACTORIAL PARAMETERS
WITH THEIR STANDARD ERRORS

Parameter	Parity 0	1	2	3	4	5
b_1	0.942 (0.005)	-0.050 (0.007)	-0.179 (0.013)	-0.050 (0.007)	-0.050 (0.007)	-0.016 (0.028)
b_2	0.006 (0.002)	0.006 (0.002)	0.006 (0.002)	0.033 (0.016)	-0.034 (0.021)	NS
b_3	NS	-0.009 (0.007)	0.045 (0.008)	NS	0.045 (0.008)	NS
b_4	NS	0.030 (0.005)	0.030 (0.005)	0.030 (0.005)	-0.030 (0.027)	0.056 (0.034)
b_5	NS	-0.010 (0.006)	-0.010 (0.006)	NS	-0.050 (0.025)	NS
b_6	NS	-0.009 (0.003)	-0.009 (0.003)	NS	-0.009 (0.003)	NS
b_7	NS	NS	NS	-0.066 (0.015)	NS	NS
b_8	NS	0.009 (0.006)	0.009 (0.006)	NS	-0.025 (0.024)	NS
b_9	NS	NS	NS	NS	-0.031 (0.020)	NS
b_{10}	NS	-0.036 (0.009)	0.051 (0.015)	NS	-0.036 (0.009)	NS
b_{11}	NS	NS	NS	NS	NS	NS
b_{12}	NS	-0.025 (0.009)	0.041 (0.015)	NS	-0.062 (0.023)	NS

TABLE 5

ANALYSIS OF VARIANCE FOR THE FINAL MODEL
FOR THE PARITY PROGRESSION RATIOS

Source of Variation	d.f.	x^2
Model	23	1198.80
Error	48	28.60

TABLE 6

PREDICTED PARITY PROGRESSION RATIOS AND MEAN NUMBER CEB
WITH THEIR STANDARD ERRORS

Sub-population			Parity 0	1	2	3	4	5	Mean
Cath	less	low	0.948 (0.006)	0.916 (0.013)	0.810 (0.020)	0.757 (0.030)	0.648 (0.043)	0.687 (0.046)	3.635 (0.146)
Cath	less	med	0.948 (0.006)	0.824 (0.023)	0.778 (0.030)	0.761 (0.032)	0.540 (0.051)	0.524 (0.057)	3.179 (0.152)
Cath	less	high	0.948 (0.006)	0.917 (0.026)	0.660 (0.042)	0.679 (0.051)	0.865 (0.070)	0.794 (0.075)	3.383 (0.263)
Cath	more	low	0.948 (0.006)	0.952 (0.013)	0.773 (0.021)	0.720 (0.033)	0.600 (0.054)	0.640 (0.059)	3.545 (0.158)
Cath	more	med	0.948 (0.006)	0.981 (0.013)	0.679 (0.030)	0.662 (0.033)	0.563 (0.048)	0.547 (0.053)	3.291 (0.127)
Cath	more	high	0.948 (0.006)	0.831 (0.022)	0.685 (0.031)	0.704 (0.037)	0.560 (0.055)	0.489 (0.066)	2.972 (0.138)
Non	less	low	0.935 (0.005)	0.909 (0.011)	0.808 (0.014)	0.822 (0.180)	0.798 (0.024)	0.838 (0.024)	3.867 (0.082)
Non	less	med	0.935 (0.005)	0.849 (0.016)	0.707 (0.022)	0.624 (0.024)	0.664 (0.037)	0.648 (0.041)	3.024 (0.097)
Non	less	high	0.935 (0.005)	0.878 (0.020)	0.727 (0.029)	0.546 (0.038)	0.644 (0.061)	0.572 (0.072)	3.009 (0.142)
Non	more	low	0.935 (0.005)	0.909 (0.012)	0.701 (0.017)	0.714 (0.023)	0.645 (0.043)	0.685 (0.046)	3.269 (0.097)
Non	more	med	0.935 (0.005)	0.869 (0.016)	0.601 (0.022)	0.517 (0.024)	0.397 (0.041)	0.381 (0.046)	2.628 (0.069)
Non	more	high	0.935 (0.005)	0.858 (0.015)	0.618 (0.021)	0.437 (0.026)	0.418 (0.044)	0.346 (0.057)	2.572 (0.062)

AN APPLICATION OF MULTIVARIATE DISCRIMINANT ANALYSIS TO PERCEPTIONS OF STUDENT PERSONNEL SERVICES

Jean M. David & Carolyn R. McIver, Florida Technological University

I. Introduction

The purpose of the research reported in this paper is to ascertain the existence of significant differences among three groups: (1) administrators; (2) faculty; and (3) students; according to their perceptions of student personnel services with respect to four areas of interest, which are: (1) importance; (2) awareness; (3) effectiveness; and (4) location. A survey has been conducted, in which the subjects have been asked questions about eight areas of student personnel services hypothesized to be important contributors to such group differences. The stepwise method of multivariate discriminant analysis was considered appropriate since this study is specifically interested in assessing the contribution of each variable to the separation of the cases into three groups. The results show that most of the variables specified do offer considerable separating power. The findings should be of interest to those whose responsibility it is to plan the allocation of currently scarce resources among student personnel services.

II. Multivariate Discriminant Analysis

In this section, the technique of multivariate discriminant analysis will be briefly described and its appropriateness for the task at hand will become clear as the discussion proceeds. The interpretation of the results of the analysis, which was conducted with the help of the sub-program discriminant (SPSS), will follow in a later section.

In multivariate discriminant analysis either the direct method or the stepwise method may be employed. The direct method is suitable only when intermediate results based on subsets of the independent variables are of no interest to the investigator. Since this study is explicitly interested in assessing the contribution of each variable, the stepwise method is deemed appropriate. This method selects independent variables for entry into the analysis on the basis of their discriminating power. Given a full set of independent variables, the sequential selection of the "next best" discriminator at each step enables a reduced set of variables to be identified, which is as good as, and sometimes better than, the full set. The process of selection begins by choosing the variable that has the highest value on the selection criterion. The discriminant criterion has been designed in such a way that it measures group-mean differences obtained after a linear combination of a set of variables has been determined with a view to getting the group means to differ widely. Since once a linear combination has been constructed, we are dealing with a single transformed variable, the F-ratio for testing the significance of the overall difference among several group means on a single variable, suggests an appropriate criterion. The test statistic is computed as follows:

$$F = \frac{SS_b \ / \ (k-1)}{SS_w \ / \ (N-k)} = \frac{SS_b \ (N - k)}{SS_w \ (k - 1)} ,$$

where k = number of groups;
N = number of individuals;
SS_b = between groups sum of squares;
SS_w = within groups sum of squares.

The variable with the highest value on the selection criterion is chosen as the initial variable for the analysis. It is next paired off with each of the other variables in turn, and the selection criterion once again computed. The new variable, which in conjunction with the initial variable produces the highest criterion value, is selected as the second variable to "enter the equation." The initial variable and the second variable are then combined with each of the remaining variables, one at a time, with a view to forming triplets which are evaluated on the criterion. The triplet with the best criterion value identifies the third variable to be selected. This procedure of locating the next variable that would yield the best criterion value, given the variables already selected, continues until all variables are selected and no additional variables provide a minimum level of improvement.

Discriminant analysis was judged appropriate for this study because it is usually employed when researchers wish to distinguish statistically between two or more groups. In order to distinguish between the groups, the researcher selects a collection of discriminating variables that measure characteristics with respect to which the groups are expected to differ. The variables selected for this study are eight categories of student personnel services with respect to their four areas of interest which are: (1) importance; (2) awareness; (3) effectiveness; and (4) location. The three groups: (1) administrators; (2) faculty; and (3) students; are compared as to their perceptions of eight categories of student personnel services which are:

1. Administration, Registration, and Records.
2. Counselling Services.
3. Financial Aid and Placement.
4. Health Services.
5. Housing and Food Services.
6. Special Services.
7. Student Activities.
8. Student Conduct.

Actually discriminant analysis goes further than ascertaining the existence of a significant difference between groups. It seeks to find one or more linear combinations of the original predictor variables that will show large differences in group means. This is equivalent to the problem of studying the direction of group differences. It is only after such linear combinations or discriminant functions have been derived, that the two research objectives of this technique may be pursued; they are: analysis and classification. It may be helpful to summarize this discussion of discriminant analysis so far before going into one or two technicalities. Discriminant analysis attempts to separate cases into groups by forming one or more linear combinations of the discriminating variables. These combinations are represented by discriminant functions of the form:

$D_i = d_{i1} Z_1 + d_{i2}Z_2 + \ldots + d_{ip}Z_p$

where D_i is the score on the discriminant function i, the d's are weighting coefficients, and the Z's are the standardized values of the p discriminating variables used in the analysis. The functions are formed in such a way as to maximize the separation of the groups. The technique also provides various tools for the interpretation of data. Such are the statistical tests for measuring the success with which the discriminating variables actually discriminate when combined into the discriminant functions. The investigator cannot proceed with the research objectives of this technique: analysis and classification, until the discriminating functions have been derived.

Observe that the maximum number of discriminating functions that can be derived is either one fewer than the number of groups or equal to the number of discriminating variables, whichever is smaller. In discriminant analysis each group (as measured by its centroid) is treated as a point, and each discriminant function is a unique (orthogonal) dimension describing the location of that group relative to the others. In any particular analysis, the last one or several functions arrived at sequentially may not contribute much of theoretical or practical importance. Such functions would consequently be statistically insignificant and may be ignored. The SPSS DISCRIMINANT subprogram used here provides two measures for assessing the importance of the latter discriminant functions. They are: the relative percentage of eigenvalues associated with the function and the canonical correlation associated with it. A brief outline of the role of eigenvalues in group differentiation may be useful at this point in our discussion.

The criterion for measuring group differentiation along the dimension specified by the vector $\underline{v}$ is the ratio λ defined below.

$$\lambda \equiv \frac{v'Bv}{v'Wv} = \frac{SS_b\ (Y)}{SS_w\ (Y)}$$ (where v', v, are (vectors and (B,W, are matrices

This follows from: if there are p predictor variables, $X_1, X_2, \ldots, X_p$, and we form a linear combination

$Y = v_1X_1 + v_2X_2 + \ldots + v_pX_p$

of these variables, the within-groups and between-groups sums-of-squares both turn out to be expressible as quadratic forms analogous to

$\Sigma y^2 = \underline{v}'S(X)\underline{v}$

This discriminant criterion was first proposed by Fisher[1] in connection with his two-group discriminant function. The criterion for group differentiation having been selected, we need to find the set of weights $[v_1, v_2, \ldots, v_p]$ that maximize the discriminant criterion. This is achieved by taking the partial derivatives of λ with respect to each of the components v_i of the vector $\underline{v}$, and setting the result equal to zero. The vector equation obtained in the process is as follows:

$$\frac{\partial\lambda}{\partial v} = \frac{2[(Bv)(v'Bv) - (v'Bv)(Wv)]}{(v'Wv)^2} = 0$$

where the last term is a pX1 null vector. Recall that $\lambda \equiv \frac{v'Bv}{v'Wv}$. Then after dividing both numerator and denominator of the middle term by v'Wv, this equation reduces to

$$\frac{2[Bv - \lambda Wv]}{v'Wv} = 0,$$

which is equivalent to

$(B - \lambda W)v = 0.$

Now premultiply both sides by W^{-1} (assuming it exists) and the following equation is obtained:

$(W^{-1}B - \lambda I)v = 0.$

This equation is of the form

$(A - \lambda I)v = 0$

and its solution will yield the eigenvalues sought. In other words the solution of $(W^{-1}B - \lambda I)v=0$, yields r non-zero eigenvalues which are denoted $\lambda_1,\lambda_2, \ldots, \lambda_r$, in descending order of magnitude, and r associated eigenvectors $v_1, v_2, \ldots, v_r$. We recall at this point that $r = \min(k - 1, p)$. Our interest in the eigenvalues λ_r lies in the following. Since the eigenvalues λ_m are, by definition, the values assumed by the discriminant criterion for linear combinations employing the elements of the corresponding eigenvectors v_m as combining weights, the eigenvector $\underline{v_1} = [v_{11}, v_{12}, \ldots, v_{1p}]$ provides a set of weights such that the transformed variable

$Y_1 = v_{11}X_1 + v_{12} X_2 + \ldots + v_{1p} X_p$

has the largest criterion, $\lambda_1 = \frac{v'_1 B_{v1}}{v'_1 W_{v1}}$,

achievable by any linear combination of the p predictor variables. It can also be shown that the remaining eigenvectors, $\underline{v_2}, \underline{v_3}, \ldots, \underline{v_r}$, have the following properties: if the elements of $\underline{v_2}$ are used as combining weights to form a second linear combination,

$Y_2 = v_{21}X_1 + v_{22}X_2 + \ldots + v_{2p} X_p$,

then Y_2 has this property: its discriminant-criterion value, λ_2, is the largest achievable by any linear combination of the X's that is uncorrelated (in the total sample) with Y_1. Similarly a third linear combination

$Y_3 = v_{31} X_1 + v_{32}X_2 + \ldots + v_{3p}X_p$

has the largest discriminant criterion value (λ_3) among all linear combinations of the X's that are uncorrelated with both Y_1 and Y_2; and so on, until Y_r, using the elements of $\underline{v_r}$ as weights, has the largest possible discriminant-criterion value among linear combinations that are uncorrelated with all preceding combinations $Y_1, Y_2, \ldots, Y_{r-1}$. The linear combinations $Y_1, Y_2, \ldots, Y_r$ are called the first, second, ..., rth linear discriminant functions differentiating optimally among the k given groups. Before presenting the computer results, we briefly review the role of canonical correlations. The discriminant criterion and canonical correlation approaches yield identical results.[2] Each discriminant-criterion value is related to the corresponding canonical-correlation/value and either value can be expressed in terms of the other. Wilks' Lambda, our main test statistic here, can be expressed in terms of either the discriminant-criterion value or the canonical-correlation value:

$$\Lambda = 1/\prod_{i=1}^{r} (1 + \lambda_i) = \prod_{i=1}^{r} (1 - \mu_i^2).$$

This statistic is inversely related to the magnitude of differences or strength of relationships.

Clearly discriminant analysis is appropriate when researchers wish to distinguish statistically between two or more groups. The predictor variables that measure characteristics with respect to which the groups are expected to differ in this study are eight categories of student personnel services with regard to four areas of interest which are: (1) importance, (2) awareness; (3) effectiveness; and (4) location.′ The three groups: (1) administrators; (2) faculty; and (3) students, are compared as to their perceptions of eight categories of student personnel services which are: (1) administration, registration, and records; (2) counselling services; (3) financial aid and placement; (4) health services; (5) housing and food services; (6) special services; (7) student activities; (8) student conduct. An interpretation of the results of the analysis reproduced in the output of subprogram DISCRIMINANT (SPSS) is presented in the next section.

III. Analysis and Interpretation of Data

The first part of the results show the selection of the discriminating variables in descending order of their respective values on the selection criterion. Only the first of the two discriminant functions subsequently derived is presented in Table 1, column (1). This is because, as revealed by the very high value (79.76) of the relative percentage of the eigenvalue associated with the first discriminant function, the second discriminant function may safely be ignored. The standardized discriminant function coefficients are used to compute the discriminant score for a case in which the original discriminating variables are in standard form. The discriminant score is computed by multiplying each discriminating variable by its corresponding coefficient and adding together these products. There is a separate score for each case on each function. Over all cases in the analysis, the score from one function will have a mean of zero and a standard deviation of one. Thus, any single score represents the number of standard deviations that case is away from the mean for all cases on the discriminant function. Should there be several discriminant functions, each case will have a score on each function. The scores for the cases within a particular group may be averaged in order to obtain the group mean on the respective function. For a single group the means on all functions are referred to as the group centroid, which is the most typical location of a case from that group in the discriminant space. A comparison of group means on each function shows how far apart the groups are along that particular dimension. The standardized discriminant function coefficients are also of analytic importance. When the sign is ignored, each coefficient represents the relative contribution of its associated variable to that function. The sign indicates whether the contribution is positive or negative. In the first discriminant function presented in Table 1, the coefficient $d_{1,1}$ = -0.0199, indicates a negative contribution to the function by the

TABLE 1

COEFFICIENTS OF FIRST[a] STANDARDIZED DISCRIMINANT FUNCTION AND OF CLASSIFICATION FUNCTIONS

Variables	First Discriminant Function	Group I Administrators	Group II Faculty	Group III Students
v. 1	-0.0199	0.08341	-0.55084	-0.50450
v. 2	-0.0297	-0.90842	-0.25280	-0.72923
v. 3	-0.0459	1.42955	2.00221	1.50159
v. 4	0.0381	0.35939	0.21167	0.66110
v. 6	0.0483	1.10254	0.83552	1.48976
v. 7	-0.0091	-0.36500	-0.66590	-0.66797
v. 8	-0.0709	0.39219	0.73798	0.06557
v. 10	-0.0857	0.33931	0.88043	-0.03817
v. 11	-0.0436	-0.23562	0.15661	-0.23321
v. 12	0.0342	0.35817	0.27522	0.63407
v. 13	0.0843	-0.47322	-0.75809	0.16455
v. 14	-0.0059	1.70331	0.06621	0.46894
v. 16	0.0326	0.10027	0.35552	0.58612
v. 17	0.0008	0.63942	0.27010	0.38416
v. 18	0.0721	-0.62087	-1.33665	-0.45678
v. 19	-0.0475	0.08759	0.06242	-0.22548
v. 20	-0.1058	0.28696	0.65234	-0.37015
v. 21	0.0527	0.39808	0.75494	1.13284
v. 22	-0.0524	0.91245	1.07064	0.35012
v. 23	0.0244	0.77175	0.75265	1.02786
v. 24	0.0934	-1.33571	-1.29473	-0.49030
v. 25	-0.0612	0.50213	0.67441	0.28214
v. 26	0.0181	0.13766	-0.22944	0.01261
v. 27	0.1036	-0.45539	-0.60567	-0.07710
v. 29	-0.0534	0.30133	0.16847	-0.14742
v. 30	-0.0088	-0.58816	0.47222	0.08977
v. 31	+0.0577	-0.40114	-0.67930	-0.13333
Constant		-35.69815	-35.95766	-35.16306

[a]Because of the relatively small contribution made by the second discriminant function, its coefficients are not reproduced here.

variable: admissions, registration, and records, with respect to the perception of importance. This may be compared with $d_{1,27}$ = 0.1036, the coefficient of: financial aid and placement with respect to the perception of location; the contribution of this variable is positive and relatively large. Five of the original variables do not appear in the discriminant functions, their contribution to the separation of the groups being of no statistical significance.

A discriminant function's associated canonical correlation also helps assess its importance. Recall that the canonical correlation squared can be interpreted as the proportion of variance in the discriminant function explained by the groups specified. Both functions are correlated with the groups. Although the first one, as we would expect, shows a higher correlation, the program provides another criterion for eliminating discriminant functions by testing for the statistical significance of discrimination not already accounted for by the earlier function. As each function is derived, starting with zero function, Wilks' Lambda is computed. Since Lambda is an inverse measure of the separating power in the original variables that has not yet been removed by the discriminant functions, the larger the value of Lambda, the smaller the amount of information remaining.

TABLE 2
EMPIRICAL TEST STATISTICS

Functions Derived	Wilks' Lambda	Chi-square	DF
0	0.2669	373.157	54
1	0.7056	99.021	26

The results in Table 2 show that after the first function had been derived, Wilks' Lambda was 0.7056. Lambda can be transformed into a chi-square statistic for any easy test of statistical significance, according to

$$V = -[N - 1 - (p + k)/2] \ln \Lambda$$
$$= -[N - 1 (p + k)/2] \sum_{m=1}^{k} \ln (1 + \lambda m)$$

where V is Bartlett's V statistic for testing the significance of an observed Λ value. This statistic is distributed approximately as a chi-square with p(k-1) degrees of freedom. Observe that the number of degrees of freedom when zero functions had been derived is given as 54 = (27)(3-1), since p = 27 predictor variables were included, and there were k = 3 groups specified. Because the successive discriminant functions are uncorrelated, the successive terms ln (1 + λm) in the second expression are statistically independent (assuming multivariate normality of the p predictor variables). Hence the additive components of V are each approximately distributed as a chi-square variate with p + k - 2m degrees of freedom (26 = 27 + 3 - 2(2)). When cumulatively subtracting V_1, V_2, and so on from V, the remainder each time is also a chi-square variate. These successive remainders become appropriate statistics for testing whether the residual discrimination after removing or "partialling out" the first discriminant function, the first and second discriminant functions, and so on, is statistically significant.

A brief comment on the classification functions provided is appropriate here. The analytic uses of discriminant analysis are complemented by its uses as a classification technique. Classification means the process by which the likely group membership of a case can be identified when the only information available is the case's values on discriminating variables. It can also be used in testing the adequacy of the derived discriminant functions. This is achieved by classifying the cases used to derive the functions in the first place, and comparing predicted group membership with actual group membership. The success of the discriminant analysis can be measured empirically by observing the proportion of correct classifications. This is shown in Table 3.

TABLE 3
PREDICTION RESULTS

Actual Group	No. of Cases	Predicted Group Membership Group I	Group II	Group III
Group I Administrators	35	24 (68.6%)	8 (22.9%)	3 (8.6%)
Group II Faculty	75	10 (13.3%)	61 (81.3%)	4 (5.3%)
Group III Students	190	6 (3.2%)	7 (3.7%)	177 (93.2%)

Percentage of "grouped" cases correctly classified = 84.33. The classification equations, one for each group, are derived from the pooled within-groups covariance matrix and the centroids for the discriminating variables. The resulting classification coefficients are to be multiplied by the raw variable values, summed together, and added onto a constant. The equations are of the form

$$C_i = c_{i1}V_1 + c_{i2}V_2 + \ldots + c_{ip}V_p + c_{io}$$

where C_i is the classification score for group i, the c_{ij}'s are the classification coefficients with c_{io} being the constant, and the V's are the raw scores on the discriminant variables. A case would be classified into the group with the highest score. The classification equations are reproduced in columns (3), (4), and (5) of Table 1.

This analysis has compared three groups as to their perceptions of four aspects of eight areas of student personnel services, which are: (1) importance; (2) awareness; (3) effectiveness; and (4) location. The results for each aspect will now be reported separately. Table 4 shows the coefficients of the first standardized discriminant function for the four areas of interest. The second discriminant functions are not reported because of the overwhelming importance of the first functions as revealed by the relative percentage of the eigenvalues associated with the first discriminant-functions for each area of interest, which were: 87.06, 85.82, 92.42, and 85.82 for importance, awareness, effectiveness and location, respectively.

TABLE 4
COEFFICIENTS OF THE FIRST STANDARDIZED DISCRIMINANT FUNCTION FOR THE FOUR AREAS OF INTEREST

Variable	Importance	Variable	Awareness
v. 1	0.21828	v. 9	-0.21745
v. 2	0.23578	v. 10	-0.11831
v. 3	0.01375	v. 11	-0.02634
v. 5	-0.30328	v. 13	0.30146
4. 6	-0.11336	v. 14	-0.16066
v. 7	-0.05864		
v. 8	-0.01609	v. 16	0.25398

Variable	Effectiveness	Variable	Location
v. 17	0.19939	v. 25	0.30913
v. 18	0.03492	v. 26	-0.01183
v. 19	-0.06641		
v. 21	-0.23548	v. 29	-0.14345
v. 22	0.21942	v. 30	0.00614
v. 23	-0.05839	v. 31	-0.28271
v. 24	-0.22946	v. 32	-0.12818

Further information is provided by the classification functions reproduced in Table 5 and the values of group centroids displayed in Table 6, while the performance of the discriminating functions may be observed from the prediction results, shown in Table 7. From the results reported it is possible to identify the composite of variables that best separate the groups.

TABLE 5
COEFFICIENTS OF CLASSIFICATION FUNCTIONS

(a) IMPORTANCE

Variable	Group I	Group II	Group III
1	-0.28489	-0.52649	-0.66491
2	-0.21609	-0.23818	-0.58235
3	1.54261	1.76657	1.66093
5	0.69420	0.81249	1.22732
6	0.63865	0.66372	0.86451
7	-0.03601	-0.28755	-0.09654
8	0.28785	0.48945	0.42869
Constant term	-23.48167	-25.50647	-25.88618
(b) AWARENESS			
9	1.34552	1.07948	0.86729
10	-0.64566	-0.55608	-0.79301
11	-0.14503	0.01582	-0.08840
13	0.64478	0.59342	1.17736
14	1.72508	1.10843	1.01942
16	0.20820	0.33207	0.63372
Constant term	-19.81915	-15.84587	-18.25594
(c) EFFECTIVENESS			
17	1.35118	1.06520	0.82743
18	-0.32646	-0.53714	-0.51488
19	-0.22755	-0.06184	-0.04429
21	0.37427	0.56954	0.96241
22	1.57449	1.43109	0.86448
23	0.71890	0.60543	0.80086
24	-0.60974	-0.38976	-0.04103
Constant term	-16.46411	-15.02886	-17.11745
(d) LOCATION			
25	0.34750	0.40405	0.15853
26	-0.19802	-0.45032	-0.36158
29	0.35721	0.39382	0.50833
30	0.32748	0.67493	0.56195
31	0.20669	-0.16279	0.25075
32	-0.18429	0.03305	0.07889
Constant term	-4.58306	-5.65010	-7.28807

TABLE 6
CENTROIDS OF GROUPS IN REDUCED SPACE
(First Functions)

Group	Importance	Awareness	Effect-iveness	Location
I	0.26139	-0.40536	0.28595	-0.12060
II	0.18188	-0.22681	0.16273	0.80180
III	-0.11994	0.16420	-0.11691	0.00979

TABLE 7
PREDICTION RESULTS

(a) IMPORTANCE

Actual Group	No. of Cases	Predicted Group Membership Group I	Group II	Group III
Group I	35	22 (62.9%)	9 (25.7%)	4 (11.4%)
Group II	75	26 (34.7%)	30 (40.0%)	19 (25.3%)
Group III	190	30 (15.8%)	53 (27.9%)	107 (56.3%)

Percentage of "Grouped" cases correctly classified: 53.00

(b) AWARENESS

Actual Group	No. of Cases	Group I	Group II	Group III
Group I	35	21 (60.0%)	6 (17.1%)	8 (22.9%)
Group II	75	11 (14.7%)	44 (58.7%)	20 (26.7%)
Group III	190	20 (10.5%)	34 (17.9%)	136 (71.6%)

Percentage of "Grouped" cases correctly classified: 67.00

(c) EFFECTIVENESS

Actual Group	No. of Cases	Group I	Group II	Group III
Group I	35	24 (68.6%)	6 (17.1%)	5 (14.3%)
Group II	75	30 (40.0%)	28 (37.3%)	17 (22.7%)
Group III	190	15 (7.9%)	30 (15.8%)	145 (76.3%)

Percentage of "Grouped" cases correctly classified: 65.67

(d) LOCATION

Actual Group	No. of Cases	Group I	Group II	Group III
Group I	35	16 (45.7%)	9 (25.7%)	10 (28.6%)
Group II	75	17 (22.7%)	39 (52.0%)	19 (25.3%)
Group III	190	22 (11.6%)	45 (23.7%)	123 (64.7%)

Percentage of "Grouped" cases correctly classified: 59.33

IV. SUMMARY OF FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

In the area of importance, three of the eight variables show a mean for the student population lower than the means for the administration and for the faculty. All three variables: (1) Admission, Registration and Records; (2) Couselling; (3) and Financial Aid and Placement; are included in the most parsimonious composite of variables separating the groups. The lower scores are indicative of perceptions of higher importance. The mean for the student population is higher than the means for the administration and faculty for both health services and special services. The mean for the student population for student activities is lower than the mean for the administration but higher than the mean for the faculty, whereas the mean for the student population for student conduct is higher than the mean for the administration but lower than the mean for the faculty. The only variable that does not contribute significantly to the separation of the groups in the area of importance is food services.

The mean for the student population is lower than the means for the administration and the faculty in the area of awareness for the same three variables as in the area of importance. The mean for the student population is higher than the means for the two other groups for the variables student conduct and health services. For three variables: food services, special services and student activities, the mean for the student population is lower than the mean for the administration but higher than that for the faculty. Food services and student activities do not make significant contributions to the separation of the groups.

The higher means for the student population for all the variables, except admission, registration, and records, where the mean for the administration is higher, indicate a lower perception of the effectiveness of these services. The lower means for all three groups than in the area of importance would indicate the services are generally considered effective. Only food services make no significant contribution to the separation of the groups in the area of effectiveness.

The means for the student population are higher than those for the other two groups for all the variables. This indicates lower perceptions of location of all these services, which is rather unexpected. Financial aid and placement, and food services are not included in the most parsimonious composite of variables separating the groups in this area.

CONCLUSIONS

The results clearly indicate that the four hypotheses that there would be no significant differences in the perceptions of importance, awareness, effectiveness, and location, respectively, of student personnel services among administrators, faculty, and students in the eight areas specified, must be rejected.

---

[1]R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, VII (1936), 179-188.

[2]Maurice M. Tatsuoka, The Relationship between Canonical and Discriminant Analysis (Cambridge, Mass.: Educational Research Corporation, 1953).

APPENDIX I

List of Original Variables

v. 1 Admissions, Registration, & Records... with respect to Importance
v. 2 Counseling " " " "
v. 3 Financial Aid & Placement " "
v. 4 Food Services " " " "
v. 5 Health Services " " " "
v. 6 Special Services " " "
v. 7 Student Activities " " "
v. 8 Student Conduct " " " "
v. 9 Admissions, Registration, & Records... with respect to Awareness
v. 10 Counseling " " " "
v. 11 Financial Aid & Placement " "
v. 12 Food Services " " " "
v. 13 Health Services..with respect to Awareness
v. 14 Special Services " " " "
v. 15 Student Activities" " " "
v. 16 Student Conduct " " " "
v. 17 Admissions, Registration, & Records... with respect to Effectiveness
v. 18 Counseling " " " "
v. 19 Financial Aid & Placement "
v. 20 Food Services" " " "
v. 21 Health Services " " "
v. 22 Special Services " " "
v. 23 Student Activities" " "
v. 24 Student Conduct " " "
v. 25 Admissions, Registration, & Records.... with respect to Location
v. 26 Counseling " " " "
v. 27 Financial Aid & Placement
v. 28 Food Services " " "
v. 29 Health Services " " "
v. 30 Special Services " " "
v. 31 Student Activities" " "
v. 32 Student Conduct " " "

APPENDIX II

The following is an example of one of the questionnaire statements and of the four perceptual response questions that will be used in this study:

Statement No. 35 - Information is available to individual students concerning all types of occupational opportunities for university graduates and requirements for these fields.

1. In your opinion, how important is this student personnel service function to a university education?
   (a) Very important__ (c) Fairly important__
   (b) Important__ (d) Not significant__
2. Are you aware of the existence of this student personnel service function on the campus at the University of Oklahome?
   (a) Highly aware__ (c) Fairly aware__
   (b) Aware__ (d) Not aware__
3. How effectively do you perceive this student personnel function on the campus at The University of Oklahoma?
   (a) Outstanding __ (c) Ineffective__
   (b) Adequate __ (d) No reaction__
4. Do you know the location where this student personnel service function is performed on the campus at The University of Oklahome?
   (a) Highly agree__ (c) Fairly agree__
   (b) Agree__ (d) Disagree__

A REVIEW OF THE APPLICABILITY OF DISCRIMINANT ANALYSIS TO SOCIAL SCIENCE RESEARCH

Beth Dawson, Southern Illinois University School of Medicine

Discriminant analysis is a statistical technique that allows the researcher to examine differences among groups. As such, it provides a unified approach to research situations where there are several groups and it is desirable to: 1) establish significant group differences; 2) study and describe the variables on which groups differ; 3) classify future individuals into the most appropriate group (Tatsouka and Tiedeman, 1954).

The primary purpose of the present paper is to describe the relevance of discriminant analysis in social science research. This will be attempted by reviewing variable selection methods, appropriate statistical tests of significance, and interpretation of discriminant functions. Finally, a brief review of some of the studies utilizing discriminant analysis in education is presented with a discussion of the relevance of the procedure for specific applications.

Both conceptually and resultwise, Tatsouka (1971) points out that discriminant analysis in the two group case is equivalent to regression analysis using a "dummy" criterion variable. More generally, with k groups and p predictor variables, results are equivalent to those obtained using canonical correlation when the set of criterion variables contains k-1 variables coded dichotomously such that:

$Y_{ij}=1$ if observation j is on a member of group i
$=0$ otherwise; $i=1,2\ldots,k-1$; $j=1,2,\ldots,n_k$.

While k-1 (or p, if smaller) discriminant functions are generated, it is almost always the case that two or three functions are sufficient to account for the group differences. Thus, the procedure results in a reduction of the space necessary to describe group differences. In addition to examining main effects, Tatsouka (1969) suggests that the method is equally effective to study the nature of interaction effects following multivariate analysis of variance. In this case, $\underset{\sim}{S}_{int}$, the sums of squares and cross products matrix (SSCP) of the appropriate interaction effect is substituted for the between groups matrix in the characteristic equation, and the eigenvalue-eigenvector pairs are found for $\underset{\sim}{W}^{-1}\underset{\sim}{S}_{int}$, where $\underset{\sim}{W}$ is the within groups SSCP matrix.

One of the practical problems confronting the researcher who uses discriminant analysis is similar to that of multiple regression: selection of variables that significantly differentiate between the groups. The problem is confounded in the case of discriminant analysis due to the additional consideration of number of groups. Several authors have addressed the variable selection problem, and, while there is no consensus on the "best" method, the following are some of the approaches that have been suggested.

In the two group case, Collier (1963) developed a test of whether the addition of p-q predictors add significantly to the discrimination obtained using q predictors alone. Assuming bivariate normality and equal variance-covariance matrices, Collier shows that

$$\frac{(n_1+n_2-p-1)}{p-q}\cdot\frac{R^2(p)\ R^2(q)}{1-R^2(p)} \sim F(p-q,n_1+n_2-p-1)$$

under the assumptions of regression where n_1 and n_2 are sample sizes of groups 1 and 2. The generalized distance function can be estimated by

$$D^2(p) = \Sigma_i^p\ \Sigma_j^p\ s^{ij}(\bar{X}_{1i}-\bar{X}_{2i})(\bar{X}_{1j}-\bar{X}_{2j})$$

where s^{ij} is the ijth element of $\underset{\sim}{S}^{-1}$, the sample estimate of $\underset{\sim}{\Sigma}^{-1}$. Using the functional relationship between R^2 and D^2, Collier suggests the following statistic for testing whether the generalized distance is the same when p variables are used as when only q variables are used ($q<p$):

$$\frac{(n_1+n_2-p-1)}{p-q}\cdot\frac{n_1n_2(D^2(p)-D^2(q))}{(n_1+n_2)(n_1+n_2-2)+n_1n_2D^2(q)}$$

distributed $F(p-q,n_1+n_2-p-1)$.

Related to tests for multivariate analysis of variance, Roy and Bargmann (1958) developed an a posteriori procedure for testing specific contributions by predictor variables. The step-down method is applied to an a priori ordering of the predictor variables: the most important variables are ordered first and the least important are ordered last. Testing occurs in reverse order and stops as soon as a significant variable is encountered.

Bargmann (1970) recommends a hierarachical method to select variables that may be used when more than one discriminant function is retained. He suggests the correlation between the first discriminant function and the predictor variables be used as an index of discrimination. Those variables with high correlations would than be deleted and the analysis redone on the remaining variables. This process is continued until most of the variables have been assigned to one of the functions in the hierarchy or until correlations become small enough to ignore.

Kraatz (1975) compares three different methods of variable selection. The first method is based upon transforming predictor variables into predictive components using the validities of these variables with a single criterion (Tucker, 1973). Kraatz extends the procedure to the special case where group membership is the criterion variable and compares this method of variable selection with a stepwise procedure and with the use of all predictor variables in a Monte Carlo study. The criterion used to judge the effectiveness of each method was the value of Wilks' lambda computed for the population. When sample values of lambda are computed for each of the three methods, the method of using all variables consistently overestimates the amount of variance accounted for between groups in the population. When population values are used for the within and total covariance matrices, all methods underestimate the amount of variance accounted for, although the method based on predictive composites of validities is the best estimator of lambda for middle ranges values of lambda. The stepwise method is never a best estimate in this situation. Finally, when a cross validation sample is considered, the predictive composite method again performs most satisfactorily.

Huberty (1971) presents a review of variable selection methods and then compares six methods. Two sets of actual data with three and five groups respectively were used for analysis and the index of discriminatory power was the proportion of correct classifications. The six selection methods were based upon beta weights, F-ratios resulting from a priori univariate analysis of variance, the stepwise procedure used in the BMD computer program, component loadings following a rotation of a principal component analysis, factor analytic-discriminatory correlations (selecting those variables that correlate highly with the first discriminant function for each of the clusters of variables defined by a factor analysis), and correlations between variables and the leading discriminant function. In general, Huberty found that the stepwise procedure results in a higher proportion of correct classifications the majority of the time.

The entire subject of variable selection needs more study. It would appear that Monte Carlo studies varying the number of variables, the number of groups and the intercorrelations between variables are in order. In addition, the effect of violation of the assumptions of multivariate normality and common variance-covariance matrices on various methods needs investigation. At the present time, there appears to be no clear-cut answer as to which variable selection method is best; it is quite possible that no one method will prove robust under all of the conditions mentioned above.

Several tests have been suggested as appropriate in determining the significance of the discrimination between groups. These tests follow naturally from the tests suggested in multivariate analysis of variance. The most commonly used statistics as given by Tatsouka (1971) are the likelihood ratio statistic, Wilks' lambda (Λ); Hotelling s trace criterion (τ); and Roy's largest root criterion (Θ).

In the case of discriminant analysis, the formulae are, letting

$r = \min(k-1,p)$
$\underset{\sim}{S}_e$ = SSCP matrix for error effects
$\underset{\sim}{S}_h$ = SSCP matrix for hypothesis effects
λ_i = the ith eigenvalue

$$\Lambda h = \frac{|\underset{\sim}{S}_e|}{|\underset{\sim}{S}_h+\underset{\sim}{S}_e|} = |\underset{\sim}{S}_e^{-1}\underset{\sim}{S}_h+\underset{\sim}{I}|^{-1} = \Pi_i^r(1+\lambda_i)^{-1}$$

$$\tau = \Sigma_i^r\lambda_i = \Sigma_i^r(\underset{\sim}{S}_e^{-1}\underset{\sim}{S}_h)_{ii} = tr(\underset{\sim}{S}_e^{-1}\ \underset{\sim}{S}_h)$$

$$\Theta = \lambda_1/(1+\lambda_1)$$

Pillai derived the distribution for Λ and tabled the null distribution for selected centile points. Heck and Pillai also developed tables for selected centile points of Θ following Roy's derivation of the null distribtuion of Θ (Bock and Haggard, 1968). The sampling distribution of Λ was obtained by Schatzoff but exact numerical computations are feasible only when p is even <u>and</u> k is odd. Consequently, two approximations <u>to</u> Λ commonly used are Bartlett's V, a function of the logarithm of Λ with an approximate chi-square distribution; and Rao's R, a radical function of Λ with an approximate F distribution. While R is a closer approximate test and is exact for p=1 <u>or</u> k=2, V an be expressed as the sum of several terms (Tatsuoka, 1971).

Schatzoff (1966) investigated these three statistics along with three others to determine their respective sensitivity to a wide variety of parameterizations. As the criterion, Schatzoff used the concept of expected significance level (ESL) that may be defined as one minus the power of a test, averaged over uniform values for α. Therefore, when comparing two test statistics, the one with the smaller ESL is preferred. Results from a Monte Carlo analysis indicate that while on one statistic is best in all situations, the largest root Θ is much worse than the other two except for the two group case. He concludes that a researcher would do equally well to use either Wilks' Λ or Hotelling's τ since both are fairly robust with respect to a wide range of alternate hypotheses.

Another approach to the question of significant discrimination is the use of a statistic such as Hays' $\hat{\omega}^2$, an estimate of the amount of variance that can be attributed to the relevant independent variable(s) in the population. Tatsuoka (1970) defined a multivariate analogue of $\hat{\omega}^2$ for use in one-way multivariate analysis of variance, denoted

$$\hat{\omega}^2_{mult} = \frac{|\underset{\sim}{S}_h| - |\underset{\sim}{S}_e| - ((k-1)/(N-k))|\underset{\sim}{S}_e|}{|\underset{\sim}{S}_h| + (1/(n-k))|\underset{\sim}{S}_e|}$$

where N is the total sample size. He performed a Monte Carlo simulation (1973a) to examine the statistical properties of $\hat{\omega}^2$mult. The statistic was found to be quite positvely biased unless the ratio of subjects to variables is extremely large. Therefore, an empirical approach to correct the bias was conducted with the result

$$\hat{\omega}^2_{corr} = \hat{\omega}^2_{mult} - \frac{p^2+(k-1)^2}{3N}\ (1-\hat{\omega}^2_{mult})$$

Upon testing, the corrected index was judged adequate for cases where $p(k-1)\leq 49$ and $75\leq N\leq 2000$. To test the significance of individual discriminant functions, Tatsuoka (1970, 1971) utilized the fact that Bartlett's V function of Λ is approximately distributed chi-square. Since the discriminant functions are uncorrelated and the terms in V are statistically independent, a sequence of tests may be defined to assess the contribution of each subsequent root. The testing procedure continues until for some j the test is not significant, leading to the conclusion that only the first j+1 discriminant functions are significant. Although subsequently questioned by Harris (1974), a recent open letter to Harris from Tatsuoka (1976) has re-established the validity of this test.

Two methods have been suggested for interpreting the relative contribution to discrimination by a given predictor variable. The first method (Tatsuoka, 1970, 1971) consists of standardizing the coefficients of the predictor variables in the discriminant function. The other method (Cooley & Lohnes, 1971) is based upon the factor structure defined by the correlations between the original variables and each of the discriminant functions. This approach is more meaningful when it is desired to interpret the discriminant functions as opposed to measuring the relative contribution of each variable to the discriminant functions (Tatsuoka, 1973b). Bargmann (1970) recommends the use of correlations between variables and only the first discriminant function.

Huberty (1975) reports a Monte Carlo study designed to study the stability of the standardized coefficients and variable-discriminant function correlations under repeated sampling. Neither measure cross validated with great stability, and the study was limited to the first discriminant function.

Following is a brief review of several studies using discriminant analysis in education. Emphasis is placed on techniques used and thoroughness of reporting rather than results of the analyses. The articles are not intended to be either representative or comprehensive; the reader is referred to the article by Tatsuoka and Tiedeman (1954) for references to uses of discriminant analysis up to that date and to the publications by Huberty (1975) and Lachenbruch (1975) for more complete listings of recent applications.

Discriminant analysis has been used in several studies related to evaluation of student success in higher education. Selover (1942), in one of the earliest applications, investigated the utility of a sophomore testing program in determining success in major curriculum groups. Data were obtained from four years of testing, and 21 curriculum groupings were analyzed. Using four subtests, Selover computed a linear discriminant function to differentiate between various sets of two groups each.

More recently, Keenen and Holmes (1970) studied college withdrawal and failure. Groups were defined on students who graduated, withdrew, or failed at a liberal arts college. Four intellectual variables (not described by the authors) and thirty non-intellectual measures based on demographic variables and stated interests were considered. Two discriminant functions were extracted on each of three analyses: the first included both types of variables, the second only intellectual variables, and the third only non-intellectual measures. Cross validation of classification using non-intellectual measures was performed on a ten percent holdout sample.

Study techniques were examined to discriminate between students in four undergraduate disciplines who had above and below average grades (Goldman & Warren, 1973). A four by two multivariate analysis of variance was performed; interaction between discipline and grade average was not significant. Subsequent discriminant analysis of the discipline effect resulted in two significant functions; no classification or cross validation analyses were reported.

Discriminant analysis has been applied in educational evaluation to study the effect of differences in learning environments. Anderson <u>et al.</u> (1969) used student perceptions of learning climates to evaluate the impact of a new physics curriculum. Three groups were defined on the basis of the experimental curriculum, the traditional curriculum, and the experience of the teachers with the new curriculum. Results of a questionnaire administered to students were the predictor variables. Discriminant analysis was followed by rotation of the two significant functions to increase interpretability. This study is uncommon with respect to the clarity and detail of reporting the methods of analysis.

A fairly common application of discriminant analysis has been the prediction of vocational choice. Cohen (1971) reports results of a study to differentiate between male college seniors in business administration and teacher education. The single discriminant function was obtained using factors from an instrument measuring motivation for career choice. Results of classification on a replication sample were reported. Porebski (1966) used three tests to discriminate between four technical trade groups. Two discriminant functions were retained; no classification was performed. Using only three predictor variables enabled the author to present considerable computational detail.

In the field of medical education, Checker <u>et al</u>. (1972) attempted to predict specialty choice for four major specialty fields from the four Medical College Admission Test (MCAT) scores obtained from students prior to admission to medical school. Two functions were retained for analysis, and classification as correct in 36 percent of the cases. No replication studies were reported.

Paiva <u>et al</u>. (1974) examined career choice transition during medical school. Three major career categories were considered. Students were surveyed as to their career preference upon entering medical school and again at the end of the senior year; the consistencies and shifts in career choice were used to define several groups. Two discriminant functions accounting for 77 percent of the differentiation were extracted and retained for interpretation. No classification analysis was performed.

Applied research in the social sciences is often hampered due to small sample sizes available for analysis. A study not so impacted resulted in a very comprehensive and detailed report on the use of discriminant analysis on Project TALENT five-year follow-up data by Hall (1967). In one study, Hall used 30 predictor variables to differentiate between 17 private colleges subsequently attended by students for whom 1960 test scores were available. In another study, Hall utilized twelve predictor variables related to student perceptions of self, home, and future and how they changed between grades nine and twelve. These papers are presented in great detail and explanation of variables used and procedures performed.

The final study to be summarized is one performed by the present author (1975) which indicates that insights may be gained into the generalizability of performance from one class to another during the developmental years of a new institution. In an exploratory study, discriminant analysis was used to identify the pre-admission variables that predict success in the first year of medical school. For the purpose of this analysis, success was defined in terms of a composite criterion score derived from performance on various examinations during the first year. Students were divided into three groups on the basis of their composite score.

Although multiple regression analysis is also an appropriate statistical procedure to predict success, the medical school curriculum is competency based and there is little desire to use a procedure that results in a precise ranking of predicted success. It is more relevant to identify those students who may experience some difficulties in the first year and for whom remedial instruction should be anticipated. Likewise, it is helpful to identify students who are expected to perform consistently above criterion level so that educational enrichment activities may be developed.

Seventeen predictor variables were examined and included the four MCAT scores, various GPA measures, the number of course hours in given areas, and data on specific courses. Data on the classes admitted in 1973 and 1974 were analyzed; sample sizes were 48 and 63. For each class, discriminant functions were found and then used to classify both the analysis class and the other class for cross validation.

A stepwise method was used to select the significant variables in the analysis. Entry of variables was controlled by using an F value of 2.00, approximating an $\alpha \leq .05$ level. The selection criterion used in this analysis was the method in which the variable that minimizes Wilks' lambda is entered at each step. Prior probabilities equal to the sample percentages were assigned for use in subsequent classification analysis.

Analysis on the first class resulted in six significant variables loading on the discriminant functions. The eigenvalues associated with the two discriminant functions account for 71.7% and 28.3% of the total variance existing in the discriminating variables (See Table I). The canonical correlations are .63 and .46 for the first and second functions respectively. Application of the sequential chi-square test of individual functions results in $p \leq .001$ for one root, $p \leq .08$ for two roots.

Table II shows a plot of the group centroids and indicates that function one primarily separates the High and Low groups, and function two differentiates between Medium and Low groups. Due to the evidence of group separation and the exploratory nature of the analysis, it was decided to retain both discriminant functions.

Interpretation of the standardized weights in Table III indicates that the most important dimension separating the groups is an ability in quantitative and science areas along with a history of specific science courses. The groups are ordered from low to high on this dimension consistent with their performance on the composite criterion. Another important dimension is verbal ability, with relevant contributions from a course in biochemistry, that separates the Medium group from the Low group. One might tentatively suggest that students well based in science with attendant ability as measured by the Quantitative MCAT perform best in the first year of medical school; students with less science and quantitative ability may still perform satisfactorily if they are well qualified on verbally related skills and have specifically taken a course in biochemistry. These results are not particularly surprising, but they tend to reinforce our prior expectations.

TABLE I

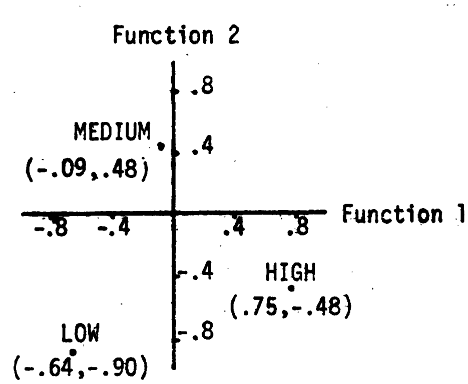
DETERMINATION OF NUMBER OF DISCRIMINANT FUNCTIONS TO BE RETAINED

Discrm Function	Eigenvalue	Relative Percentage	Canonical Correlation
1	0.670	71.69	0.633
2	0.265	28.31	0.457

Function Derived	Wilks' Lambda	Chi-Square	df	Significance
0	0.474	31.78	12	.001
1	0.791	9.98	5	.076

TABLE II

PLOT OF GROUP CENTROIDS



The final step in this analysis was to examine the accuracy of classification of students into correct groups. Comparison of actual group membership with predicted membership is shown in Table IV; 72.9% of the students are correctly classified using two functions. Computing these functions on the second class for cross validation results in correct classification of 56.6%.

Although complete details are not presented here, it is of interest to note some of the results obtained when the discriminant analysis was performed on the second class. Again, six variables loaded on the discriminant functions and both functions were retained. The functions correctly classified 84.9% of the class on whom they were computed and 54.2% of the cross validation class.

Only two variables entering the discriminant functions were the same both both analyses, an indication that the two classes may differ in abilities that tend to discriminate performance. Alternately, the composite score used as a criterion may not be an equivalent measure for the two classes. In both analyses, a considerable reduction in the percent of correct classifications occurs when the functions are applied to the cross validation sample. Although some shrinkage in correct classification is not unusual, differences in the classes or inequivalence of the criterion may be contributing to this shrinkage. Thus, the present analysis has identified two areas for further research and indicates that generalization with respect to performance of students from one class to another should be extremely cautious in the formative years of a curriculum.

TABLE III

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	Function 1	Function 2
Chemistry GPA	.322	-.277
Biochemistry Taken	.217	.651
Cell Biology Taken	.309	.101
Verbal MCAT	-.236	.626
Quantitative MCAT	.593	-.045
All Other GPA	-.072	1.016

GROUP CENTROIDS

	Function 1	Function 2
High Group	.75	-.48
Medium Group	-.09	.48
Low Group	-.64	-.90

TABLE IV

RESULTS OF CLASSIFICATION

Class on Whom Analysis Performed

Actual Group	No. of Cases	Predicted Group High	Med	Low
High	11	6	4	1
Med	28	2	25	1
Low	9	0	5	4

Percent Correctly Classified: 72.9%

Second (Validation) Class

Actual Group	No. of Cases	Predicted Group High	Med	Low
High	15	3	12	0
Med	28	2	22	4
Low	10	0	5	5

Percent Correctly Classified: 56.6%

Summary

The foregoing discussion has attempted to identify some of the methodological areas of research in discriminant analysis and examine the applicability of this method in a few selected studies. Social science research almost invariable attends to situations that are complex to study. Multiple independent and dependent measures are frequently assessed. Researchers in these fields have recently begun to utilize the greater analytic power made possible by multivariate methods. Discriminant analysis appears to be especially relevant to many research questions that involve a polychotomous, unranked criterion. The concurrent sophistication of statistical computer programs has rendered multivariate techniques accessible and easy to apply. Perhaps they are too easy; researchers should be aware of the assumptions of the models they utilize and remain cognizant of the measurement problems still existing in the study of human behaviors.

In reading reports of research that have used discriminant analysis, two particular aspects were noted that, if properly attended, could result in better research and more relevant utilization of research findings. First is the problem in discriminant analysis of the validity of the initial classification of subjects into groups. The model assumes that classification partitions the sample: each individual is a member of one and only one group. The validity and generalizability of all subsequent analyses are based upon the accuracy of the initial classification. This calls for more precise definition of groups being studied. In many situations, definition of group membership is simple and presents no problem; i.e., classification on the basis of sex, age, other biographic characteristics. Studies that examine career choice or college major must recognize that the classification is valid only for that point in time and does not necessarily extend to future preferences. The problem is even more apparent in research that utilizes diagnostic categories to describe groups, such as medical research and personality studies. The accuracy of the initial classification must be examined and findings interpreted accordingly.

Second, although it may seem elementary, greater attention should be given to complete and precise description of the research. The very nature of the groups analyzed, the basis for determining initial classifications, and the types of predictor variables used to derive the discriminant functions are frequently referred to in a very cursory manner. In addition, the lack of description of the exact procedures used is sometimes appalling. Method(s) of variable selection, criteria used to determine the number of functions retained for analysis, and the type of classification functions used should be explicitly noted in all reports of research. The space constraints of many professional journals make complete description difficult. But a brief paragraph outlining methods and criteria should be insisted upon by the author; such description need not be eloquent and can generally be condensed into several concise sentences. Certainly the primary purpose of publishing research results is to communicate methods and findings to other researchers. Many published studies do not achieve this objective.

References

Anderson, G.J et al. Curriculum effects on the social climate of learning; A new representation of discriminant functions. Amer. Educ. Research J., 1969, 6:315-328.

Bargmann, R.E. Interpretation and use of a generalized discriminant function. In R.E. Bose, et al. (Eds.), Essays in Probability and Statistics. Chapel Hill: Univ. of N.C. Press, 1970.

Bock, R.D. et al. The use of multivariate analysis of variance in behavior research. In D.K. Whitla (Ed.) Handbook of Measurement and Assessment in Behavioral Sciences. Reading, MS: Addison-Wesley, 1968.

Checker, A. et al. Can the choice of medical specialty be predicted by the Medical College Admission Test? Proceedings, 80th Annual Convention, APA, 1972.

Cohen, D. Differentiating motivations underlying vocational choice. J. of Educ. Research, 1971, 64:229-234.

Collier, R.O. A note on the multiple regression techniques for deleting variables in the discriminant function. J. of Exper. Educ., 1963, 31:351-353.

Dawson, B.K. Discriminant analysis. Paper presented at AAMC/RIME, Washington, D.C., 1975.

Goldman, R.D. et al. Discriminant analysis of study strategies connected with college grade success in different major fields. J. of Educ. Meas., 1973, 10:39-47.

Hall, C.E. Three papers in multivariate analysis. Project TALENT Five-Year Follow-up studies: Interim Report 2, AIR and Univ. of Pittsburg, 1967.

Harris, R.J. Multivariate analysis of variance. Paper presented at AERA, Chicago, 1974.

Huberty, C.J. On the variable selection problem in multiple group discriminant analysis. Paper presented at AERA, New York, 1971.

Huberty, C.J. Discriminant analysis. Review of Educ. Research, 1975, 45:543-598.

Keenen, C.B. et al. Predicting graduation withdrawal and failure in college by multiple discriminant analysis. J. of Educ. Meas., 1970, 7:91-95.

Kraatz, J. The use of predictive composites in multiple discriminant analysis. Unpubl. doct. dissertation, Urbana: Univ. of Illinois, 1975.

Lachenbruch, P.A. Discriminant Analysis. New York: Hafner, 1975.

Paiva, R.E.A. et al. Career choice transition during medical school. Paper presented at AAMC Colloq. on Career Dev't of Physicians, Washington, D.C., 1974.

Porebski, O.R. Discriminatory and canonical analysis of technical college data. Brit. J. of Math and Stat. Psych., 1966, 19:215-236.

Roy, S.N. et al. Tests of multiple independence and the associated confidence bounds. Annals of Math. Stat., 1958, 29:491-503.

Schatzoff, M. Sensitivity comparisons among tests of the general linear hypothesis. Amer. Stat. Assn. J., 1966, 415-435.

Selover, R.B. A study of the sophomore testing program at the University of Minnesota. J. of Applied Psych., 1942, 26:296-307, 456-467, 587-593.

Tatsuoka, M.M. Multivariate Analysis. Review of Educ. Research, 1969, 39:739-744.

Tatsuoka, M.M. Discriminant Analysis. Champaign, IL: Institute for Personality and Ability Testing, 1970.

Tatsuoka, M.M. Multivariate Analysis. New York: Wiley, 1971.

Tatsuoka, M.M. An examination of the statistical properties of a multivariate measure of strength of relationship. Final Report, Proj. No. 2-E-020, HEW, 1973a.

Tatsuoka, M.M. Multivariate analysis in educational research. In F.N. Kerlinger (Ed.), Review of Research in Education. Itaska, IL: Peacock, 1973b.

Tatsuoka, M.M. Letter of response to Richard Harris. Personal communication, 1976.

Tatsuoka, M.M. et al. Discriminant analysis. Review of Educ. Research, 1954, 24:402-420.

Tucker, L.R. Notes on a system to determine predictive composites based on validities. Unpubl. paper (mimeo), Urbana: Univ. of Illinois, 1973.

ON THE THEORY OF RANDOMIZED RESPONSES TO TWO SENSITIVE QUESTIONS*

Wanzer Drane, Southern Methodist University

This paper sets forth a stochastic model which can be applied to randomized response designs wherein more than one question is sensitive to stigmatizing. Using three specific designs applied to a pair of stigmatizing or sensitive traits leading to two by two contingency tables, it is proved that if the two traits are statistically independent then the responses are also and conversely. The converse statement gives the support we need for inference to be made on the traits based on the randomized responses.

INTRODUCTION

Let Π be a vector of *trait* probabilities which are unknown; let Λ be a corresponding vector of *response* probabilities; and let P be a nonsingular *randomizing* matrix. By definition we set $\Lambda = P\Pi$. Since Λ is estimated directly from the responses, $\hat{\Pi} = P^{-1}\hat{\Lambda}$ and $\mathrm{Cov}(\hat{\Pi}) = P^{-1}\Sigma P^{-1'}$ wherein $\Sigma = \mathrm{Cov}(\hat{\Lambda})$.

This differs from the Warner (1971) approach in that his P matrix is the expectation of a stochastic matrix, and P is singular "requiring" weighted least squares to estimate Π. In the present model P is a transition matrix between two state vectors represented by Π and Λ, respectively, the trait and response state probabilities. An element of P, p_{ij}, is the probability of obtaining the response i, given a subject is in state j.

Finally, Π and Λ are vectors of cell probabilities of contingency tables. Using the initial Warner design (1965), the Simmons or Unrelated Question design (Greenberg, et al. 1969) and the Forced Yes design (Drane, 1975) it is proved that "if the two (stigmatizing) traits are independent, then the responses are also and conversely."

It is the converse that is essential to proper inferences being drawn from analyses of contingency tables constructed as the result of a randomized response design survey.

1. A STOCHASTIC MODEL FOR TWO SENSITIVE QUESTIONS

Let us consider two sensitive or stigmatizing traits and two randomized questions, each directed at only one of these traits. For the purposes of this treatment the questions and their respective randomizing devices will give rise to YES-NO answers, only. Presuming independence between trials and a random selection of individuals, each respondent produces an entry to a 2×2 contingency table. If π_j are the trait probabilities, λ_i the response probabilities and p_{ij} the conditional probability that a person in state j will produce a response i, then the randomized response can be summarized as

$$\Lambda = P\Pi \tag{1.1}$$

or, by resubscripting according to 2×2 tables,

$$\begin{bmatrix}\lambda_{11}\\ \lambda_{12}\\ \lambda_{21}\\ \lambda_{22}\end{bmatrix} = P\begin{bmatrix}\pi_{11}\\ \pi_{12}\\ \pi_{21}\\ \pi_{22}\end{bmatrix} \tag{1.2}$$

wherein P is a 4×4 randomizing or transition matrix. Furthermore we shall require P to be nonsingular. The 2×2 Trait Table will be that of Figure 1 with one for one substitutions when considering the corresponding 2×2 Table.

		Trait 2 Y	Trait 2 N	
Trait 1	Y	π_{11}	π_{12}	$\pi_{1.}$
	N	π_{21}	π_{22}	$\pi_{2.}$
		$\pi_{.1}$	$\pi_{.2}$	1

FIG. 1

Table of Trait Probabilities

In the figure, π_{11} = probability a person drawn at random possesses both traits, π_{12} = probability that he possesses the first but not the second, etc. The $\pi_{i.}$ and $\pi_{.j}$ are the corresponding marginal probabilities. The table of response probabilities is obtained by substituting λ everywhere for π and "Response" for "Trait."

The requirement that P be nonsingular results from the fact that inferences on Π are made from observations made with probabilities Λ. Thus, if information about Λ, say $\hat{\Lambda}$, is to be preserved with no loss whatever, then not only should (1.1) and (1.2) hold, but

$$\Pi = P^{-1}\Lambda \tag{1.3}$$

should also be true, so that

$$\hat{\Pi} = P^{-1}\hat{\Lambda} \tag{1.4}$$

which is a restatement of the invariance principle of maximum likelihood estimates.

The usual hypothesis for independence is that Π can be written as a vector of products, namely

$$\Pi = \begin{pmatrix}\pi_{11}\\ \pi_{12}\\ \pi_{21}\\ \pi_{22}\end{pmatrix} = \begin{pmatrix}\pi_{1.} & \pi_{.1}\\ \pi_{1.} & \pi_{.2}\\ \pi_{2.} & \pi_{.1}\\ \pi_{2.} & \pi_{.2}\end{pmatrix} \tag{1.5}$$

If the two stigmatizing traits are independent, then (1.5) will be true. Once again, however, we are not observing responses to direct questions. So the appropriate question is the following: If the responses are independent, are the traits also

*Reprinted from Communications in Statistics-Theory and Methods, A5(6), 565-574 (1976), by courtesy of Marcel Dekker, Inc.

independent? The following theorem will be proved for three different designs previously proposed and/or used.

Theorem: If the two (stigmatizing) traits are independent, then the responses are also and conversely.

2. THREE DESIGNS

2.1. Forced Yes Design

The first design is a simplified version of the Morton design (see Greenberg, Horvitz and Abernathy, 1974, and Drane, 1975) in which the alternative question always has the same positive answer "Yes" or the person being questioned is required to answer "Yes".

Definitions:

p_i = probability of being asked the ith sensitive question, $i = 1, 2$,

q_i = probability of being instructed to answer "Yes" on the ith question and

π_{ij}, $\pi_{.j}$ and $\pi_{i.}$ are the trait probabilities of the 2×2 table of Figure 1.

The response table, with the same form as that of Figure 1, has entries λ_{ij} as follows:

$$\lambda_{11} = p_1p_2\pi_{11} + p_1q_2\pi_{1.} + q_1p_2\pi_{.1} + q_1q_2\pi_{..}\ , \tag{2.1}$$

where $p_1p_2\pi_{11}$ is the probability that the subject will be asked both stigmatizing questions and that the answer to both will be "Yes;" $p_1q_2\pi_{1.}$ is the probability that only the first stigmatizing question will be asked and the response will be "Yes", forcing a "yes" instead of the second question; $q_1p_2\pi_{.1}$ is the reverse; and $q_1q_2\pi_{..}$ is the probability that each answer will be a forced yes and $\pi_{..} = \pi_{11} + \pi_{12} + \pi_{21} + \pi_{22} = 1$.

Substituting for $\pi_{1.}$, $\pi_{.1}$ and $\pi_{..}$ we get,

$$\lambda_{11} = \pi_{11} + q_2\pi_{12} + q_1\pi_{21} + q_1q_2\pi_{22}. \tag{2.2}$$

In a manner similar to that used in deriving (2.1) and (2.2) we obtain the matrix equation (1.1) wherein

$$P = \begin{pmatrix} 1 & q_2 & q_1 & q_1q_2 \\ 0 & p_2 & 0 & q_1p_2 \\ 0 & 0 & p_1 & p_1q_2 \\ 0 & 0 & 0 & p_1p_2 \end{pmatrix}, \tag{2.3}$$

and

$$P^{-1} = \frac{1}{p_1p_2}\begin{pmatrix} p_1p_2 & -p_1q_2 & -q_1p_2 & q_1q_2 \\ 0 & p_1 & 0 & -q_1 \\ 0 & 0 & p_2 & -q_2 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.4}$$

If we apply the hypothesis that $\pi_{ij} = \pi_{i.}\pi_{.j}$ and substitute this and (2.3) into (1.1), we obtain

$$\Lambda = \begin{pmatrix} (\pi_{1.} + q_1\pi_{2.})(\pi_{.1} + q_2\pi_{.2}) \\ (\pi_{1.} + q_1\pi_{2.})p_2\pi_{.2} \\ p_1\pi_{2.}(\pi_{.1} + q_2\pi_{.2}) \\ p_1\pi_{2.}p_2\pi_{.2} \end{pmatrix}, \tag{2.5}$$

which proves the direct theorem. The converse theorem is proved by substituting (2.4) into (1.3) and assuming, $\lambda_{ij} = \lambda_{i.}\lambda_{.j}$, independence among the responses. In this case we obtain

$$\Pi = \frac{1}{p_1p_2}\begin{pmatrix} (p_1\lambda_{1.} - q_1\lambda_{2.})(p_2\lambda_{.1} - q_2\lambda_{.2}) \\ (p_1\lambda_{1.} - q_1\lambda_{2.})\lambda_{.2} \\ \lambda_{2.}(p_2\lambda_{.1} - q_2\lambda_{.2}) \\ \lambda_{2.}\lambda_{.2} \end{pmatrix} \tag{2.6}$$

which proves the converse theorem.

2.2. Initial Warner Design

Instead of the alternative being a forced Yes, Warner (1965) proposed an alternate question or statement which was the negative of the sensitive question. The respondent was asked to verify (Yes, No) one of the following: "I belong to Group A," or "I do not belong to Group A," with probabilities p_1 and q_1 respectively. We extend that design here to two questions and ask for verifications of "I belong to Group B" or "I do not belong to Group B" with probabilities p_2 and q_2. The trait and response probabilities, π_{ij} and λ_{ij}, are as before stated. Applying the step by step procedure as in the preceding example, we obtain

$$P = \begin{pmatrix} p_1p_2 & p_1q_2 & q_1p_2 & q_1q_2 \\ p_1q_2 & p_1p_2 & q_1q_2 & q_1p_2 \\ q_1p_2 & q_1q_2 & p_1p_2 & p_1q_2 \\ q_1q_2 & q_1p_2 & p_1q_2 & p_1p_2 \end{pmatrix}, \tag{2.7}$$

and

$$P^{-1} = \frac{1}{(p_1-q_1)(p_2-q_2)}\begin{pmatrix} p_1p_2 & -p_1q_2 & -q_1p_2 & q_1q_2 \\ -p_1q_2 & p_1p_2 & q_1q_2 & -q_1p_2 \\ -q_1p_2 & q_1q_2 & p_1p_2 & -p_1q_2 \\ q_1q_2 & -q_1p_2 & -p_1q_2 & p_1p_2 \end{pmatrix}. \tag{2.8}$$

Once again applying the hypothesis of independence to Π and Λ, (2.7) and (2.8) inserted into (1.1) and (1.3), respectively, yield

$$\Lambda = \begin{pmatrix} (p_1\pi_{1.} + q_1\pi_{2.})(p_2\pi_{.1} + q_2\pi_{.2}) \\ (p_1\pi_{1.} + q_1\pi_{2.})(q_2\pi_{.1} + p_2\pi_{.2}) \\ (q_1\pi_{1.} + p_1\pi_{2.})(p_2\pi_{.1} + q_2\pi_{.2}) \\ (q_1\pi_{1.} + p_1\pi_{2.})(q_2\pi_{.1} + p_2\pi_{.2}) \end{pmatrix} \tag{2.9}$$

and

$$\Pi = \frac{1}{(p_1-q_1)(p_2-q_2)}\begin{pmatrix}(p_1\lambda_{1.} - q_1\lambda_{2.})(p_2\lambda_{.1} - q_2\lambda_{.2}) \\ (p_1\lambda_{1.} - q_1\lambda_{2.})(-q_2\lambda_{.1} + p_2\lambda_{.2}) \\ (q_1\lambda_{1.} - p_1\lambda_{2.})(-p_2\lambda_{.1} + q_2\lambda_{.2}) \\ (q_1\lambda_{1.} - p_1\lambda_{2.})(q_2\lambda_{.1} - p_2\lambda_{.2})\end{pmatrix} \tag{2.10}$$

which are proofs of both the theorem and its converse.

2.3. Unrelated Question Model

Walt Simmons suggested (Greenberg et al. 1969) that the second statement of Warner be replaced by an innocuous statement which was a priori unrelated to the stigmatizing statement, and which innocuous trait occurred with known probability. We again let

p_i = probability of being asked the ith sensitive question,
q_i = probability of being asked the ith innocuous question,
ϕ_i = probability of "Yes" to the ith innocuous question, and
$\theta_i = 1 - \phi_i$ (i = 1, 2 in all cases),

with Π and Λ as before. A given respondent may give up any combination of answers regardless of his true trait state. But the randomizing matrix is beginning to show some complexity.

P =

$$\begin{pmatrix}(p_1+q_1\phi_1)(p_2+q_2\phi_2) & (p_1+q_1\phi_1)q_2\phi_2 & q_1\phi_1(p_2+q_2\phi_2) & q_1\phi_1q_2\phi_2 \\ (p_1+q_1\phi_1)q_2\phi_2 & (p_1+q_1\phi_1)(p_2+q_2\phi_2) & q_1\phi_1q_2\phi_2 & q_1\phi_1(p_2+q_2\phi_2) \\ q_1\phi_1(p_2+q_2\phi_2) & q_1\phi_1q_2\phi_2 & (p_1+q_1\phi_1)(p_2+q_2\phi_2) & (p_1+q_1\phi_1)q_2\phi_2 \\ q_1\phi_1q_2\phi_2 & q_1\phi_1(p_2+q_2\phi_2) & (p_1+q_1\phi_1)q_2\phi_2 & (p_1+q_1\phi_1)(p_2+q_2\phi_2)\end{pmatrix} \tag{2.11}$$

and

$P^{-1} = 1/p_1p_2$

$$\cdot\begin{pmatrix}(p_1+q_1\phi_1)(p_2+q_2\phi_2) & -(p_1+q_1\phi_1)q_2\phi_2 & -q_1\phi_1(p_2+q_2\phi_2) & q_1\phi_1q_2\phi_2 \\ -(p_1+q_1\phi_1)q_2\phi_2 & (p_1+q_1\phi_1)(p_2+q_2\phi_2) & q_1\phi_1q_2\phi_2 & -q_1\phi_1(p_2+q_2\phi_2) \\ -q_1\phi_1(p_2+q_2\phi_2) & q_1\phi_1q_2\phi_2 & (p_1+q_1\phi_1)(p_2+q_2\phi_2) & -(p_1+q_1\phi_1)q_2\phi_2 \\ q_1\phi_1q_2\phi_2 & -q_1\phi_1(p_2+q_2\phi_2) & -(p_1+q_1\phi_1)q_2\phi_2 & (p_1+q_1\phi_1)(p_2+q_2\phi_2)\end{pmatrix} \tag{2.12}$$

Finally, if $\pi_{ij} = \pi_{i.}\pi_{.j}$ and if $\lambda_{ij} = \lambda_{i.}\lambda_{.j}$, i, j = 1, 2, then

$$\Lambda = \begin{pmatrix}[(p_1+q_1\phi_1)\pi_{1.} + q_1\phi_1\pi_{2.}][(p_2+q_2\phi_2)\pi_{.1} + q_2\phi_2\pi_{.2}] \\ [(p_1+q_1\phi_1)\pi_{1.} + q_1\phi_1\pi_{2.}][q_2\theta_2\pi_{.1} + (p_2+q_2\theta_2)\pi_{.2}] \\ [q_1\theta_1\pi_{1.} + (p_1+q_1\theta_1)\pi_{2.}][(p_2+q_2\phi_2)\pi_{.1} + q_2\phi_2\pi_{.2}] \\ [q_1\theta_1\pi_{1.} + (p_1+q_1\theta_1)\pi_{2.}][q_2\theta_2\pi_{.1} + (p_2+q_2\theta_2)\pi_{.2}]\end{pmatrix} \tag{2.13}$$

and

$$\Pi = \frac{1}{p_1p_2}\begin{pmatrix}[(p_1+q_1\theta_1)\lambda_{1.} - q_1\phi_1\lambda_{2.}][(p_2+q_2\theta_2)\lambda_{.1} - q_2\phi_2\lambda_{.2}] \\ [-(p_1+q_1\theta_1)\lambda_{1.} + q_1\phi_1\lambda_{2.}][q_2\theta_2\lambda_{.1} - (p_2+q_2\phi_2)\lambda_{.2}] \\ [-q_1\theta_1\lambda_{1.} + (p_1+q_1\phi_1)\lambda_{2.}][(p_2+q_2\theta_2)\lambda_{.1} - q_2\phi_2\lambda_{.2}] \\ [q_1\theta_1\lambda_{1.} - (p_1+q_1\phi_1)\lambda_{2.}][q_2\theta_2\lambda_{.1} - (p_2+q_2\phi_2\lambda_{.2}]\end{pmatrix}, \tag{2.14}$$

which proves both the theorem and its converse for the Unrelated Question design.

3. SUMMARY

For all three designs independence between either traits or responses implies the other. The randomizing matrix P contains this information. The columns of P always add to one regardless of whether independence is preserved between Π and Λ. It is a requirement that the rows of P be the coefficients of a binomial product of the form

$$\lambda_{ij} = (a_i\pi_{1.} + b_i\pi_{2.})(c_i\pi_{.1} + d_i\pi_{.2}) \tag{3.1}$$

if independence between the traits imply independence between the responses. At this writing it is not known what is required of P so that Π also has row elements of the form of (3.1) with λ and π interchanged.

It is easy to see that the Unrelated Question design of Simmons reduces to the Forced Yes design by setting $\phi_i = 1$ and $\theta_i = 0$. The determinant of P in both cases is $(p_1p_2)^2$.

The requirement that $p_i \neq q_i$ in the Warner design is again in evidence. Neither of the other two designs reduces to this one by manipulating ϕ and θ.

Many other designs are left untouched by this paper, as is the connection between efficiency, power, etc. and P. These are left as objects of further investigations.

This paper has established that inferences regarding independence between the responses in a 2×2 table with either of the foregoing designs carries over to the traits as well.

Remarks:

Bill Parr, a student at Southern Methodist University, has pointed out that on purely logical grounds we have the following proposition: If independence among the traits implies independence among responses, then dependence among responses implies dependence among the traits. So that inferring dependence among responses carries over in this case, but it is possible without the above theorem to have independence among responses and dependence among traits.

Gould, Shah and Abernathy (1969) developed a behavioral model which fits into a 2×2 table and has 42 parameters in it which has to be some kind of a record. Then without asking questions of independence, as it was built on repeated sampling, they estimated as many as five parameters the rest being accounted for by constraints.

Lastly, bibliographies can be found in two expository articles by Greenberg, Horvitz and Abernathy (1974) and Horvitz, Greenberg and Abernathy (1975).

BIBLIOGRAPHY

[1] Drane, W. (1975). Randomized response to more than one question. Proc. Social Sci. Sect. Amer. Statist. Assoc., 395-7.
[2] Gould, A. L., Shah, B. V., & Abernathy, J. R. (1969). Unrelated question randomized response techniques with two trials per respondent. Proc. Social Sci. Sect. Amer. Statist. Assoc., 351-9.

[3] Greenberg, B. G., Abul-Ela, A. A.-L., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model theoretical framework. J. Amer. Statist. Assoc., 64, 520-9.

[4] Greenberg, B. G., Horvitz, D. G., & Abernathy, J. R. (1974). A comparison of randomized response designs. Reliability and Biometry (F. Proschan & R. J. Serfling, Eds.). Philadelphia: Society for Industrial and Applied Mathematics, 787-815.

[5] Horvitz, D. G., Greenberg, B. G., & Abernathy, J. R. (1975). Recent developments in randomized response designs. A Survey of Statistical Design and Linear Models (J. N. Srivastava, Ed.). New York: American Elsevier Publishing Co., Inc., 271-85.

[6] Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. J. Amer. Statist. Assoc., 60, 63-9.

[7] Warner, S. L. (1971). The Linear randomized response model. J. Amer. Statist. Assoc., 66, 884-8.

ASSESSING UNCERTAINTY OF PREDICTION WHEN ADDITIONAL INDEPENDENT VARIABLES ARE CONSIDERED FOR REGRESSION

Eugene Dutoit, US Army Infantry Center and Columbus College
Douglas Penfield, Rutgers University

I. Introduction.

The purpose of this paper is to discuss another rule for determining whether the addition of another independent variable into a linear multiple regression equation is worth the effort required to obtain data on the variable. In addition, the paper attempts to give a simple explanation of why the standard error of estimate about the regression surface may increase when additional variables are entered.

The paper is divided into three parts. Part I is the introduction, Part II examines the algebraic relationships and Part III discusses applications and gives an example using data.

II. Development.

The form of the linear multiple regression equation considered in this study is:

$$y = b_o + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k \quad (1)$$

The common measure of "goodness of fit" is the coefficient of determination (R^2_k) which is computed as the ratio of explained variance (by regression) to the total variation. Tables have been constructed and published that show what the critical value of R_k would have to be if "k" variables were included in the multiple regression equation. Consider the relationship for computing the standard error of estimate:

$$S^2_{y \downarrow X_k} = \frac{n-1}{n-k-1} (S^2_y)(1-R^2_k) \quad (2)$$

If the analyst has the option to include k, k+1, k+2, . . . number of independent variables in his prediction equation he may want some measure of the change in uncertainty of prediction about the regression. One measure of uncertainty about prediction is the standard error of estimate ($S_{y \downarrow X_k}$) - (Equation 2). Consider the case where the number of observations is fixed. The standard errors of estimate for k and (k+1) independent variables would be $S_{y \downarrow X_k}$ and $S_{y \downarrow X_{k+1}}$ respectively.

It is a common conception that the inclusion of additional variables in the regression equation always tends to reduce the error of estimate somewhat and leads to an increase in R. From the above statement it can be inferred that:

$$S_{y \downarrow X_k} > S_{y \downarrow X_{k+1}} \quad (3)$$

The following ratio can be computed:

$$\frac{S_{y \downarrow X_k}}{S_{y \downarrow X_{k+1}}} \quad (4)$$

This may be interpreted as a ratio of the standard error of estimates computed for (k) and (k+1) independent variables. The importance of this ratio is developed below and illustrated in Figure 1.

Inequality (3) is based on the common conception stated above. However, the underlined statement is not true for all combinations of values for n, k, R^2_k and R^2_{k+1}. The following discussion explores the total relationship in more depth. Consider equation (2) and its extension denoted by (2A).

$$S^2_{y \downarrow X_{k+1}} = \left(\frac{n-1}{n-(k+1)-1}\right) (S^2_y)(1-R^2_{k+1}) \quad (2A)$$

Equation (2A) can be written as:

$$S^2_{y \downarrow X_{k+1}} = \left(\frac{n-1}{n-k-2}\right) (S^2_y)(1-R^2_{k+1}) \quad (2A)$$

When one more independent variable is added to the regression function it follows that:

$$\frac{n-1}{n-k-2} > \frac{n-1}{n-k-1} \quad (5)$$

It is also true that:

$$R^2_{k+1} > R^2_k \quad (6)$$

for an additional independent variable. From equation (6) it then follows that:

$$(1-R^2_k) > (1-R^2_{k+1}) \quad (7)$$

Both (n-1) and (S^2_y) remain the same in equations (2) and (2A). Ratio (4) may be expanded through equations (2) and (2A).

$$\frac{S^2_{y \downarrow X_k}}{S^2_{y \downarrow X_{k+1}}} = \left(\frac{n-k-2}{n-k-1}\right) \cdot \left(\frac{1-R^2_k}{1-R^2_{k+1}}\right) \quad (8)$$

Where:

$$\left.\begin{array}{l} \dfrac{(n-k-2)}{(n-k-1)} < 1.0 \\ \text{and} \\ \dfrac{(1-R^2_k)}{(1-R^2_{k+1})} > 1.0 \end{array}\right\} \quad (9)$$

Therefore, the value of the ratio ($S^2_{y \downarrow X_k} / S^2_{y \downarrow X_{k+1}}$), as given in (8), will depend on how much one factor is less than one in relation to how much the other factor is greater than one--i.e., equations (9).

Equation (8) can be plotted as shown in the following figure.

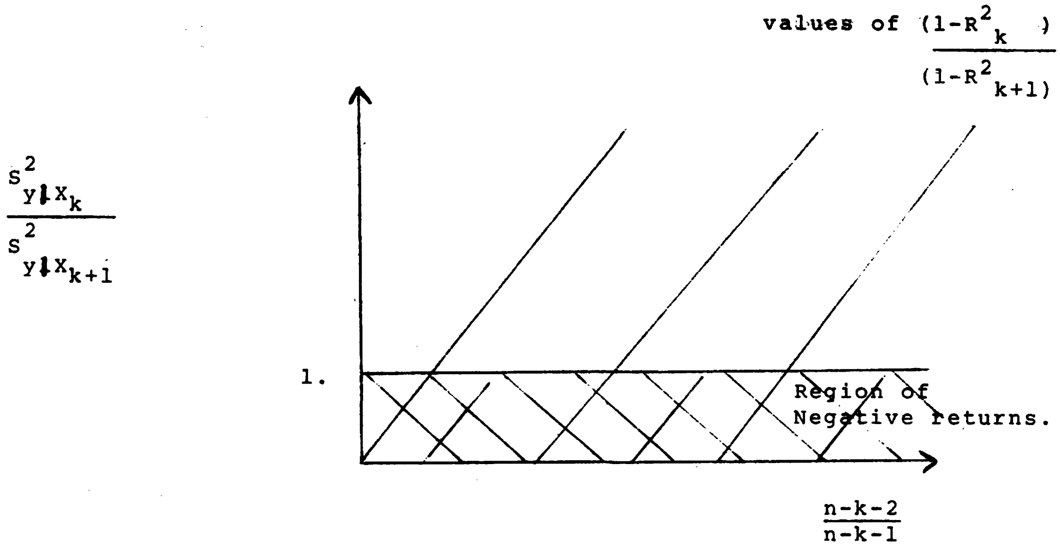


Figure 1

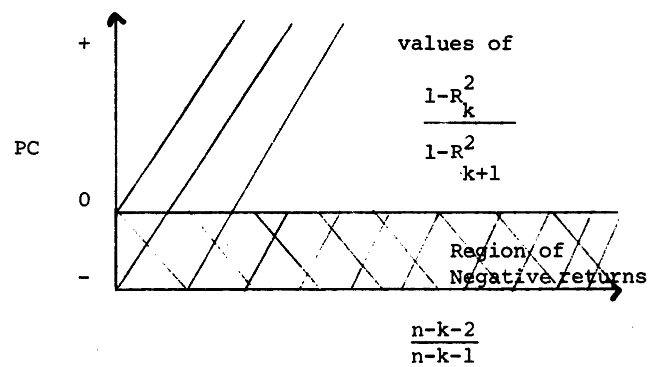


Figure (2)

Values characteristic of Figure (1) may be used by the social science researcher to determine if the relationship expressed on inequality (3) is true for his particular experimental situation. If relationship (3) is true; then, the ratio expressed as the ordinate ($s^2_{y \cdot x_k} / s^2_{y \cdot x_{k+1}}$) will be greater than one. If relationship (3) is not true, the ratio will be less than one and will fall in the shaded or region of negative returns (i.e., the value of $s^2_{y \cdot x_k} < s^2_{y \cdot x_{k+1}}$).

If inequality (3) is expected to be true, then these relationships can be modified to give more information to the analyst. The percentage change (PC) in the standard error of estimate is:

$$PC = \left[\frac{s_{y \cdot x_k} - s_{y \cdot x_{k+1}}}{s_{y \cdot x_k}}\right] \times 100 \qquad (10)$$

Equation (10) can be interpreted as some estimate percentage decrease in $s_{y \cdot x_k}$ which results by adding one more additional independent variable. Substituting equations (2) into equation (10):

$$PC = \frac{\left[\frac{n-1}{n-k-1} \cdot (1-R^2_k)\right]^{\frac{1}{2}} - \left[\frac{n-1}{n-k-2} \cdot (1-R^2_{k+1})\right]^{\frac{1}{2}}}{\left[\frac{n-1}{n-k-1} \quad (1-R^2_k)\right]^{\frac{1}{2}}} \times 100$$

which is equivalent to:

$$PC = \left[1 - \sqrt{\frac{\frac{n-1}{n-k-2} \cdot (1-R^2_{k+1})}{\frac{n-1}{n-k-1} \cdot (1-R^2_k)}}\right] \times 100$$

or finally

$$PC = \left[1 - \sqrt{\frac{n-k-1}{n-k-2} \cdot \frac{(1-R^2_{k+})}{(1-R^2_k)}}\right] \times 100 \qquad (11)$$

Equation (11) can be displayed in the same fashion as equation (8). The diagram is shown below:

Figure (2) is computed in table form for selected values of $[(n-k-2)/(n-k-1),\ R^2_k,\ R^2_{k+1}]$ - see Table 1. It should be noted that the negative percent reduction indicates that the standard error of estimate actually increases when some additional independent variables are included in the regression equation.

Table 1. Expected Percent Reduction In The Standard Error Of Estimate

$(1-R^2_k)/(1-R^2_{k+1})$

$\frac{n-k-2}{n-k-1}$	1.00	1.02	1.05	1.10	1.15	1.20
.90	-5.41	-4.37	-2.87	- .50	1.71	3.78
.92	-4.26	-3.23	-1.75	.60	2.78	4.83
.94	-3.14	-2.13	- .66	1.66	3.82	5.85
.96	-2.06	-1.06	.40	2.69	4.83	6.83
.98	-1.02	- .02	1.42	3.69	5.80	7.79
1.00	0.00	.99	2.41	4.65	6.75	8.71

III. Application.

A methodology for applying these results is outlined below:

1. For a given set of data of n observations on k independent variables the value of R_k^2 is known.

2. Perhaps the value of R_{k+1}^2 (for an additional independent variable) may be estimated within the context of the experimental situation. In some circumstances the researcher may have some idea of his estimates of R_{k+1}^2 which are based on past experience in observing the relationships between the variables. In any case, the researcher would probably want to consider a range of values (estimates) for R_{k+1}^2.

3. The values of (R_{k+1}^2) are applied in equation (11). If the estimated values of PC do not fall in the region of negative returns (i.e., the region of negative returns = $PC < 0$), then the standard error of estimate would probably be reduced if the (k+1) independent variable were included in the regression. Of course, a value of $PC > 0$ is no firm reason for selecting additional independent variables for regression. The incremental differences $D=(R_{k+1}^2 - R_k^2)$ should also be statistically significant for some value if the type I error equal to α.

Numerical Example:

The following example is taken from Johnson (1950). It consists of a random sample of 50 students taken from a study dealing with the prediction of achievement of freshmen at the University of Minnesota. The variables that were selected were:

Y = honor-point ratio of the end of the freshman year.

X_1 = score on the Johnson Science Application Test.

X_2 = score on an English Test.

X_3 = score on the Cooperative Algebra Test.

X_4 = percentile rank in high school graduation class transformed to probits ($\mu=5, \sigma=1$).

The total set of data for all 50 students can be obtained from Johnson (1950). The following lines summarize the output results of these data by using the program BMD02R.

Step 1
Variable entered = X_4
n = 50
R^2 = .2188
St'd error of estimate = .4848

Step 2
Variable entered = X_1
n = 50
R^2 = .2638
St'd error of estimate = .4757

Step 3
Variable entered = X_3
n = 50
R^2 = .2656
St'd error of estimate = .4802

Step 4
Variable entered = X_2
n = 50
R^2 = .2664
St'd error of estimate = .4853

Equation (11) is verified by example by referring to the four steps of the stepwise regression exercise shown above. Equation (11) is repeated below:

$$PC = \left[1 - \sqrt{\frac{n-k-1}{n-k-2} \cdot \frac{(1-R_{k+1}^2)}{(1-R_k^2)}} \right] X\ 100 \quad . . . \quad (11)$$

1. Step 1 to step 2

1	2
n = 50	n = 50
k = 1	k+1 = 2
R_k^2 = .2188	R_{k+1}^2 = .2638
$S_{y \cdot x_1}$ = .4848	$S_{y \cdot x_2}$ = .4757

Actual Change in $S_{y \cdot x}$:

$$\left(\frac{.4848 - .4757}{.4848}\right) \cdot 100 = 1.9\%$$

Using Eq (11)

$$\left[1 - \sqrt{\frac{(48)}{(47)} \frac{(.7362)}{(.7812)}}\right] \cdot 100 = 1.9\%$$

2. Step 2 to step 3

2	3
n = 50	n = 50
k = 2	k+1 = 3
R_k^2 = .2638	R_{k+1}^2 = .2656
$S_{y \cdot x_2}$ = .4757	$S_{y \cdot x_3}$ = .4802

Actual Change in $S_{y \cdot x}$:

$$\left(\frac{.4757 - .4802}{.4757}\right) X\ 100 = -1\%$$

Using Eq (11)

$$\left[1 - \sqrt{\frac{(47)}{(46)} \frac{(.7344)}{(.7362)}}\right] 100 = -1\%$$

3. Step 3 to step 4

3	4
n = 50	n = 50
k = 3	k+1 = 4
R_k^2 = .2656	R_{k+1}^2 = .2664
$S_{y \cdot x_3}$ = .4802	$S_{y \cdot x_4}$ = .4853

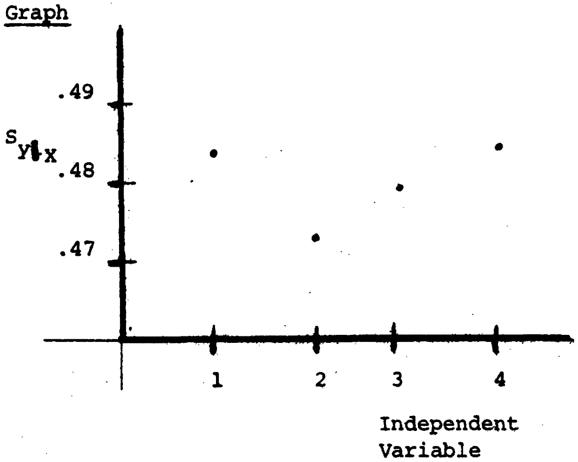
Actual Change in $S_{y \cdot x}$:

$$\left(\frac{.4802 - .4853}{.4802}\right) X\ 100 = -1.1\%$$

Using Eq (11)

$$\left[1 - \sqrt{\frac{(46)}{(45)}\frac{(.7336)}{(.7344)}}\right] \cdot 100 = -1.1\%$$

Graph



References:

BMD, Biomedical Computer Programs, Berkeley, California: University of California Press, 1970

Crow, E. and Davis, F. and Maxfield, M. Statistics Manual. New York: Dover Publications, 1960

Johnson, P. Statistical Methods in Research. New York: Pentice-Hall, Inc., 1950

A SIMULTANEOUS EQUATION MODEL OF THE U. S. CARPET INDUSTRY

Carl L. Dyer, North Carolina State University

Introduction

The carpet and rug industry of the United States grew at approximately 14% per year during the period from the late 1950's to the early 1970's, with the major portion of the growth attributable to tufted carpets and rugs. The latter constitute about 92% of the total market today, with woven carpets making up about 2.5 to 3%, needlepunch carpets accounting for about 4% and the remainder of the market consisting of knitted, braided and hooked fabrics.

The industry is predominately a tufted industry today. Value added by domestic carpet and rug manufacturers amounted to almost a billion dollars in 1973, and the dollar value of production at the mill level grew rapidly during the 1960's to exceed $3 billion in 1973, as compared to a total of less than one-fifth of that in 1960. The growth of the tufted sector ranked third among four digit SIC industries during the 1958-1969 period, based on DOC value-of-shipments data. This rate of growth was exceeded only by the semi-conductor and ammunition industry. Perhaps this outstanding growth was a contributing factor for the industry being unprepared for economic slump of 1970 and poorly prepared for the sharp downturn of 1974, but, whatever the reasons, a better information system is required for the industry and its suppliers.

In addition to basic information for decision-making at the firm level for carpet and associated firms, several governmental agencies and units could well benefit from a consistent, scientific tool for policy formation and for forecasts of sales and capacity. Such a tool, properly developed and utilized, could well assist in structuring the industry, and in so doing, provide insights into the industry that would not generally be available otherwise.

Reasons for the Study

There are multiple reasons that justify a study of this nature due to both the forecasting and the explanatory capabilities of such a model. From the forecasting aspect, a reason is that carpet manufacturers need information regarding future consumption and price levels of carpets at the time of evaluating investment alternatives. Typically, in the industry, firms note their present share of the market and then make use of long term industry forecasts as to revenue and yardage to decide on the capital appropriations to make under the assumption that they will maintain that share of the market. If they expect their share to change, they will adjust the appropriations accordingly. These "long term" forecasts, however, have typically been made with methods that are best suited for forecasting in the short term; methods such as exponential smoothing, trend and seasonal analysis are short term methods since they implicitly assume a stable and recurrent pattern of determination of the variables that exogenously have an effect on the sector or industry of interest, such as the carpet industry. For example, these short term methods would assume a stable and repeated pattern in the construction industry. If there is a change in the pattern of determination of the number of housing starts, perhaps due to a change in governmental policy with regards to subsidized housing construction, such methods would have no means of incorporating this change or of estimating what the quantitative effects of such a policy may be on the carpet industry's output.

Suppliers to the carpet industry, such as fiber producers and yarn manufacturers, would also benefit from the long-term forecasting capabilities of such a model, as they too have to make capital expenditure decisions, and in the case of the man-made fiber industry these decisions have to be made well in advance due to the capital intensiveness of the production process.

Government also has use for an econometric model of the industry; with the increased employment in carpet manufacture and with the additional consideration that a substantial portion of the industry's output is produced in one country, government needs the forecasts of this model in order to plan housing policies so as to provide an adequate spatial distribution of the labor force.

The explanatory aspects of the model are also of interest to the possible users; the model is set up in terms of structural relationships that are the demand and supply equations of carpets and of their inputs. Manufacturers of carpets are interested in the ceteris paribus effect of a change in price on the quantity demanded of carpeting of the kinds considered in this study. The same interest holds true for suppliers to the industry and in particular synthetic fiber manufacturers as carpet producers and their largest clients and the former are interested in estimating the expected change in fiber utilization due to a change in the unit price of fibers. A correctly specified model will allow for the estimation of these price elasticities. An impact multiplier, which refers to the size of the change in an endogenous variable due to a small change in the value of an exogenous variable, is also of direct policy interest. In a simplified two equation model, for example

$$Y_1 = \beta_1 Y_2 + \alpha_1 x_1 + \mu_1$$

and

$$Y_1 = \beta_2 Y_2 + \alpha_2 x_2 + \mu_2$$

where Y_1 and Y_2 are endogenous, x_1 and x_2 are exogenous, μ_1 and μ_2 are random variables with zero expected value and finite variance, and β_1, β_2, α_1 and α_2 are parameters. Solving Y_1 and Y_2:

$$Y_1 = -\frac{\alpha_1 \beta_2}{(\beta_1 - \beta_2)} x_1 + \frac{\beta_1 \alpha_2}{(\beta_1 - \beta_2)} x_2$$

$$- \frac{\beta_2}{(\beta_1 - \beta_2)} \mu_1 + \frac{\beta_1}{(\beta_1 - \beta_2)} \mu_2$$

$$Y_2 = -\frac{\alpha_1}{(\beta_1 - \beta_2)} x_1 + \frac{\alpha_1}{(\beta_1 - \beta_2)} x_2$$

$$- \frac{1}{(\beta_1 - \beta_2)} \mu_1 + \frac{1}{(\beta_1 - \beta_2)} \mu_2$$

The impact multiplier on Y_1 of a change in x_1 is $-\alpha_1\beta_2/(\beta_1-\beta_2)$, while the impact of the same change on Y_2 is $\alpha_1/(\beta_2-\beta_1)$. In reference to Y_1, the short term multiplier refers to the size of the change in Y_1 due to a unit change in x_1, whereupon x_1 will return to its original value. The long term multiplier refers to the change in the level of Y_1 due to a permanent change in the level of x_1. For example, government may consider the implementation of quotas and/or tariffs of man-made fiber products. These tariffs and/or quotas will have an effect on the prices and availability of carpet yarn, which in turn will have an effect on prices, output, and employment in the carpet industry. Government may also consider a permanent change in its policy with regards to subsidizing housing construction, or the Federal Reserve may change the availability of money, which in turn is expected to have an effect on construction starts and the growth of income. Government policies that influence the output of the automobile and the airplane industries also affect the carpet industry, though in a smaller way. Carpet manufacturers and their suppliers are interested in the quantitative effect of such policies and policy changes on the prices and output of their product. The estimation of a value for these impact multipliers is central in answering such questions.

Broadly speaking, the market for carpeting should be viewed in perspective as the market for one product which is closely interrelated with the market for other products, all of which are substitutes to a degree, that fulfill the demand for the general product floor covering. Carpeting competes with resilient flooring, hardwood flooring, and hard surface flooring to satisfy the need for floor covering in the utility function of consumers. Recent figures show that carpeting has increased its share of the floor covering market to an estimated 60% in 1975.

Two developments in the supply side of the market are at the heart of this increased usage of carpeting as the most popular flooring material; the coming about of tufting as the predominant method of manufacture, and the development of synthetic fibers which in turn were used as the face fiber in the carpet. Those two developments lowered the unit cost of production, thereby shifting the supply curve to the right, which in an environment of competition meant declining carpet prices. Substitution and income effects clearly took place, and consumption of carpeting increased.

Carpets and rugs can be classified according to multiplicity of criteria. A carpet is any floor covering material that is directly fastened to the floor and usually covers the entire floor area (from "wall to wall"), while a rug is not fastened to the floor and does not cover the entire floor area. Carpets and rugs were originally manufactured in the U. S. by some sort of weaving method such as Axminster, Wilton, or Velvet. These weaving processes are labor intensive processes where the weavers are specialized in their trade, with long training periods before their being able to weave efficiently. However, in early 1950's the tufted method of manufacture was introduced. This production process requires less skill on the part of labor, as well as being a much faster process. Thus, due to its lowered unit costs of production and its acceptability by consumers, this method of manufacture proliferated to the point that today in excess of 94% of all carpeting sold is tufted.

Carpeting is different sizes. The term broadloom refers to the carpet manufactured in long rolls - 150' to 300' - of widths 6' or more. In excess of 84% of all carpeting manufactured in 1973 was manufactured as broadloom, with about 12% manufactured as rugs 4' x 6' or smaller, and the remaining 4% manufactured in the standard 4' widths of the auto and aereo market.

Synthetic fibers have taken a commanding lead over natural fibers, with expectations that this lead will continue to increase. For example, in 1965 wool, which is the chief natural fiber used in carpets, constituted 18% of the total face fiber poundage used, with this percentage consistently declining to about 1.2% in 1975. Of the man-made fibers consumed, nylon ranks first with about 71% of the total face fiber usage, and polyester is a distant second with about 12%, and acrylic (plus modacrylic) fibers make up about 7.7% of total carpet face yarns.

The industry is structured along the lines of five sectors: the carpet, yarn, backing, labor and fiber sectors; it is explicitly presumed that changes in value of the variables that exogenously affect a variable determined in one of the sectors will bring about changes in the values of the endogenous variables.

Time Considerations

The objective of this study is to estimate the parameters of the demand and supply functions of carpets and rugs and their inputs. One of the first considerations to resolve is whether the model's parameters are to be estimated from either cross-sectional or time series data or a combination of both. As a practical matter, limitations as to the availability of the data usually resolve this question, as was partially the case in this study. A complete set of cross-sectional data was not available. Some assumptions could be made about the probability that the carpet bought in state i would be consumed in state j, but these assumptions would add to the variability of the disturbance terms of the demand relationships. Thus, due mostly to data limitations, estimation from cross-sectional data was not feasible.

A study based upon annual time series observations is not desirable because an adequate number of observations would require the study to reach back into the early fifties, and there have been severe changes in the method of manufacture and the pattern of ownership of the firms in the industry since then. Tufting, which revolutionized the industry, established itself as the principal method of manufacturing in the late fifties and early sixties, displacing the traditional weaving method. In addition, there has been a steady trend in the industry for the ownership to evolve from family-owned into publicly-held corporations. These two phenomena can be expected to have had a significant effect on the effect on the structure of the industry, and there are no adequate ways to account for these changes in the specification of the model, given the nature of available data. Furthermore, no monthly data is available for the input variables that are expected to be jointly distributed with the output variables. Consequently, the model specification was formulated on a quarterly basis, since adequate data of this type existed.

A simultaneous determination of the endogenous variables is implied. This assumption is a reasonable one due to the nature of the production process and the methods of distribution of carpeting and its inputs. There are not long lags between the decision to produce and the delivery of the product to the market as is typically the case of agricultural commodities. In these commodities, supply in the current period is frequently specified as a function of price in the previous period, while the price in the current period is determined by demand alone. In the carpet industry information on input prices is current, so decisions on current production is made with knowledge of current input prices. In addition, with the time period being a quarter of a year, it is expected that utilization of inputs during the quarter for the production of output during the same quarter depends on current prices of inputs and of outputs; that is, if there are any lags, they are much shorter than a quarter. Thus, if there is a change in the value of an exogenous variable such as imports of man-made fiber products, it is expected that this change will affect price and availability of man-made staple which in turn will affect (in the same period) price and availability of yarn which in turn will affect (again in the same time period) the price and the availability of carpet products; this is the simultaneous nature of the determination of the endogenous variables.

Nature of the Model

The model constructed in this study is of a simultaneous nature, since in the U.S. carpet industry the production process is such that typically there are no long lags between the time when a decision on production is made and its execution occurs, or when a raw material is purchased and when it is consumed in production. Thus it is reasonable to make a priori assumption that quantities and prices of outputs and of inputs are jointly distributed. That is, for a given time period, prices or quantities of outputs are determinants in the demands for inputs, but prices or quantities of inputs are determinants in the supply of outputs, and prices or quantities of one input may be determinants in the demand or supply for other inputs. The nature of the joint distribution is in addition to the usual expectation about the joint distribution of quantities and prices of a particular output (input) due to the interaction between the demand and the supply of the output (input). It is asserted that due to the simultaneous interaction of the model's relationships, an estimating procedure that accounts for these interactions must be used.

In algebraic terms, the model may be written as follows:

$$BY_t + GZ_t = U_t$$

where,

Y_t = column vector of observation of P jointly dependent variables at time t,

Z_t = column vector of observations of K predetermined variables (exogenous and lagged endogenous) at time t,

B, G = PxP and PxK matrices, respectively, of coefficients to be estimated, and

U_t = column vector of observations of P random variables at time t.

The model consists of P structural relationships to be identified that together interact to determine the values of P endogenous variables of interest. In this model, parameters of the structural equations are estimated, and, moveover, the reduced form of the model established is of interest. To obtain this, we observe that if B is non-singular, $B^{-1}BY_t + B^{-1}GZ_t = B^{-1}U_t$, or $Y_t = \pi Z_t + V_t$ is the reduced form of the model where $\pi = -B^{-1}G$, and $V_t = B^{-1}U_t$.

In the reduced form the endogenous variables are expressed explicitly in terms of predetermined variables and estimated coefficients. Thus, it is possible for one to inquire as to the quantitative effect of a change in the value of an exogenous variable on the endogenous variables, and additionally, to predict values of the endogenous variables for given, assumed, or forecasted values of the exogenous variables.

In this study, the endogenous variables of interest are prices and shipments of carpets and rugs and their inputs. The carpet and rug products for which demand and supply relationships are studied are broadloom tufted, broadloom woven, auto and aero carpeting, needle-punched, and 4' x 6' and smaller, while the inputs are fiber, yarn, backing, and labor. Capital is not explicitly considered an input, and this study is essentially short run in nature, where the short run refers to a decision period (of time) in which the capital of the firms is fixed, but other inputs may vary in quantity. In these studies capital is taken as given, and in every time period the static position of the firms is observed and their utilization of the variable inputs recorded.

The empirical model consists of 16 equations and 60 variables, of which 27 are endogenous, 21 are exogenous and 12 are predetermined. Estimation was by two-stage least squares and ordinary least squares.

The Carpet Market

The carpet market sector consists of 12 equations; five demand, five supply, one inventory, and one advertising equation. These latter two are not an integral part of the carpet market per se, but are included with this sector because one is related to the distribution of carpets, while the other is related to their demand; thus, they fit better in this sector than in any of the other sectors. The carpet products under consideration are: tufted broadloom, woven broadloom, needlepunched, and other, auto and aereo, and 4' x 6' and smaller.

The Yarn Market

The yarn market is represented by three equations, one equilibrium condition, and a statement about one of the supply conditions. Spun yarn and filament yarn are competitive inputs and substitution of one for the other occurs to some degree in the carpet yarn market. Additionally, there are differences in the method of distribution that suggest the two yarns be treated separately. Filament yarn on the one hand is sold directly by the fiber producers to carpet manufacturers, whereas spun yarn goes through an additional stage, the spinning stage. Staple is typically sold by fiber producers to yarn spinners who in turn sell to carpet manufacturers, except for yarn spun by carpet manufacturers. There is heavy competition among yarn spinners who react very readily to short run market conditions. This is not the case of the more stable pricing policies of the fiber manufacturers.

It is argued that typically the supply of filament yarn is perfectly elastic which is tantamount to saying that at the predetermined price, consumers of filament yarn can purchase all of the yarn they desire to purchase. This is a consequence of the conditions of excess capacity that have characterized the fiber industry during the period of analysis.

The Backing Market

The two types of backing which are considered in this market are the jute backing and the man-made type. The supply of man-made backing is considered to be perfectly elastic. There has generally been ample supply of this backing during the period of this study. Thus, the price of synthetic backing is considered as predetermined. With jute, the situation is not as clear-cut. Jute is imported from India, Pakistan, and Bangladesh, which accounts for about 85% of the world's jute production. Therefore, the supply of jute is intimately related to the political developments in those countries. The political conditions were highly unstable during much of the time period bridged by this study, so much so that the supply of jute seems to have alternated between perfect elasticity and perfect inelasticity, which dictates that a reasonable assumption be made. In times of peace, the supply can be assumed to have been perfectly elastic as there is ample supply of jute and labor in those countries, whereas in times of war, the supply can be assumed to have been perfectly inelastic as there was a breakdown of the climate conducive to exporting. U. S. dock strikes have also played a role. For purposes of estimation, we assume that the most typical situation during the period was that of a perfectly elastic supply curve.

The Labor Market

The market for this input is characterized by a predetermined quantity supplied. It has been pointed out by several people that the Dalton area in northwest Georgia produces in excess of 55% of the carpet manufactured, but a shortage of appropriate housing limits the size of the labor force in this area. However, in the production of woven carpeting, the skills necessary to be a carpet weaver require long training periods, so the number of weavers is assumed in the model to be given. Thus, in the labor market quantity will be predetermined, and wage adjustments are allowed to bring the market into equilibrium.

The Fiber Market

One of the aspects of a simultaneous equation model is that it allows for the joint distribution of variables, which is a more realistic and adequate representation of real world behavior. Specifically, in this case, it allows for the joint distribution of quantities of fibers. Thus, it is possible to follow the quantitative effect of a change on a variable that affects prices and quantities of fiber, on prices and quantities of carpet. The fiber sector is represented by one equation, the derived demand equation (derived from the derived demand for spun yarn) for fiber and a statement about supply being perfectly elastic for the reasons of excess capacity outlined in the description of the yarn market. The price of cellulosic staple is included as there is some substitution between noncellulosic and cellulosic fiber in the manufacture of broadwoven goods; this substitution is especially true in the case of blends such as rayon-polyester and others. The quantity of broadwoven goods is considered as being exogenously determined due to income and other characteristics which are of interest in a study of the textile industry, but not in one of the carpet industry. Finally, imports of synthetic fiber products are expected to be inversely related with noncellulosic staple's quantity demanded.

A Stock Adjustment

A stock adjustment model where producers move towards the quantity they want to supply, but because of capital restrictions are unable to completely reach, is more characteristic of the carpet industry's supply response than is the traditional formulation. Therefore, the supply, the inventory, and the advertising equations are specified as a compact set of equations where the only explanatory variables are the ones judged to have a substantial impact on supply. This model is compatible with the assumptions of profit maximization.

The stock adjustment model is based on the hypothesis that at a given price of the output, producers desire to supply a given unobservable quantity; in the context of the model as used in this study, it is a rate adjustment model, which by definition is quantity per unit of time. At the given price, producers desire to bring to the market a certain quantity of carpet per unit of time; for instance, a quarter. Producers desire to adjust the rates (not the stocks) at which they desire to bring the product to the market. Assume a linear relationship

$$Q_1^* = \beta_o + \beta_1 P_t$$

where,

Q_t^* = desired supply during period t,

P_t = price of the output during period t,

β_o, β_1 = unknown parameters.

If the supply during period t-1 was Q_{t-1}, producers due to equipment and manpower limitations, can change production by a fraction of the desired change:

$$(Q_t - Q_{t-1}) = \gamma (Q_t^* - Q_{t-1}),$$

where,

$0 < \gamma < 1$, and Q_t = actual supply in period t.

This fractional change works in the downward as well as in the upward direction since producers are reluctant to rid themselves of labor which they would have to retrain. This fractional flexibility in the downward direction is especially true in the carpet industry where a tight supply of labor is the rule. It might be noted that this hypothesized behavior of manufacturers where current supply is a function of lagged supply is more realistic as it implies smoothed production rates. Substituting Q_t^* from the first equation into the second yields:

$$Q_t = \alpha_o + \alpha_1 P_t + (1 - \gamma) Q_{t-1}$$

where, $\alpha_o = \gamma\beta_o$ and $\alpha_1 = \gamma\beta_1$.

It can be argued that α_1 is the price coefficient of the short run supply, while β_1 is the price coefficient of the long run supply. This assertion seems to place a restriction on the long run supply in the form that its price coefficient must be of the same sign as the one in the short run, which precludes the possibility of increasing returns to scale in the long run. Besides, time series studies are typically interpreted as short run in nature where the rates of utilization of inputs and of output generation are observed for given prices of the inputs and of the output, thereby making it questionable whether long run inferences can be drawn from observing these short run rates. Perhaps what can be said is that α_1 is a better estimator of the short run price parameter than the coefficient estimated by regressing Q_t directly on P_t.

In statistical estimation of cause and effect relationships, it is necessary first that there be recurrent autonomous forces operating to determine the values of the variables. If these forces do not have an autonomy of their own, it is doubtful whether one can measure them successfully. The preliminary estimates of the needle-punched demand and supply price parameter results seemed to bear this out, and thus they were omitted from the model.

In addition to the stock adjustments, a seasonal variable was added as a regressor in the demand equations. The interest in econometric estimation in this study is to measure the net effect, for example, of income on quantity demanded of a particular carpet product. Therefore, the purely seasonal effect of carpet has to be removed to leave the net effect of income.

Results and Summary

Microeconomic theory was used as a foundation upon which to specify a model of the carpet and rug industry. The model constructed consisted primarily of demand and supply equations for various carpet products and their inputs, where the various relationships interacted to jointly determine prices, quantities, and other endogenous variables. These structural equations were then estimated through the use of econometric methods, and inferences were drawn. Finally, the structure was solved for the endogenous variables, which would facilitate forecasting values of the endogenous variables, although this was not done in this study. Due to space limitations, the equations of the model and the empirical results are not presented here, but a general summary is provided.

The model was estimated in two blocks. In the first block, the 14 equations of the carpet, yarn, backing, labor, and fiber markets were estimated with two-stage least squares estimators, while in the second block the inventory and advertising equations were estimated with ordinary least squares estimators. The latter estimation was possible since the levels of inventory and advertising were postulated as functions of predetermined variables alone and the error terms of each equation were assumed to be independent of one another and of the error terms of the equations of the variables in the first block.

Broadly speaking, the results in the estimated structure serve to reinforce the a priori beliefs about the industry's operation. For example, the demand function for tufted broadloom was found to be highly price elastic over the relevant price range, and thus the equilibrium quantity tended to be sensitive to shifts in supply. In addition, industry output was found to be very sensitive to prices of synthetic fibers. The carpet industry has historically been in the fortunate position of utilizing the general decline in fiber prices and thus owes its growth in part to this decline. The effect of this decline was to shift the supply of carpets out and to the right, thus intersecting the relatively flat demand curve at higher quantities than if the demand were relatively steep. The further change in fiber costs will play a significant role in the growth of the industry.

An inference drawn from the model estimates indicated that advertising as viewed from the point of view of the industry as a whole is not a critical variable in determining carpet consumption. Interfirm allocations are a different matter but it is doubtful that even then one firm would sell significantly more than another due to more advertising alone. In other words, advertising does not move carpets as well as price does. The apparent fact that price moves carpets may explain in part the declining trend in the number of pounds of yarn per square yard of carpet that exists in the industry. By "thinning out" the product, carpet manufacturers have been able to lower prices and thus reach a wider market.

Another interesting inference drawn from the model is found in the relationship between residential and commercial markets. If tufted broadloom sales for residential construction (housing starts) and sales from nonresidential construction (office building, etc.) are picking up the contract market, which they largely were until late 1974, then it is interesting to note that the elasticity of demand of nonresidential construction was about twice the elasticity of demand of residential construction. Hence, the carpet industry apparently is more sensitive to relative and percentage in nonresidential than in residential construction.

Another empirical finding that reinforces a priori beliefs deals with the price cross-elasticity of demand for tufted broadloom and for woven broadloom. The demand for tufted broadloom is more sensitive to changes in woven prices than is the demand for woven broadloom to changes in tufted prices. Some consumers readily shift in and out of the tufted market depending, ceteris paribus, on what woven prices are doing. If woven prices were to decrease enough, a group of consumers will readily switch to woven; if woven prices became too high, these same consumers would return to the tufted market. If tufted prices decrease, ceteris paribus, proportionally fewer consumers will switch away from woven carpets, and if tufted prices increase, fewer consumers will switch into woven; they may well withdraw from the market for carpeting together if the earlier conclusion that price was a very important variable in moving carpets is uniquely true.

Finally, this study is only a beginning stage in the development of a full model of the textile industry of the U. S. Work in underway by the author to develop models for other sectors of the industry, and refinement of this model of the carpet industry will continue.

RURAL-URBAN DIFFERENTIALS IN MISSISSIPPI WHITE-NONWHITE EMPLOYMENT: 1950-1970*

M. E. El Attar
Mississippi State University

INTRODUCTION

Urban growth or rural farm atrophy is a product of fundamental socio-economic changes and technological developments. The major process of urbanization in Mississippi has begun since 1940 and is still occuring today. The size of urban population in 1970 (987,312 or 44.5 percent) is more than twice that for 1940 (432,882 or 19.8 percent). On the other hand, there has been a sizable population concentration in rural non-farm areas. The rural non-farm population has roughly tripled its share to the total population, increasing from 16.6 percent (351,030) in 1940 to 46.0 percent (1,019,277) in 1970. The rural farm areas are the only residential types which continue to incur population losses, decreasing from 64.1 percent (1,399,884) in 1940 to 9.5 percent (210,323) in 1970 (El Attar, 1974: 6). In the light of this spatial population redistribution, it is of sociological interest to examine and compare the rural-urban differentials in peoples and jobs in the state. Mississippi has traditionally been viewed as a rural state whose urban areas do not differ from the rural localities with regard to either the occupational or industrial structure of its employed residents.

Rural-urban differentials in the occupational and industrial structures of employed residents are phenomena which have relatively received rather limited attention during this redistribution period. Quite understandably, sociological and demographic research in Mississippi has focused on rural areas. As a consequence, we know comparatively little about the occupational and industrial structures of urban residents in comparison to those of their rural counterparts. It is in this type of differentiation that the goals of the present paper are identified.

DATA, DEFINITIONS, AND SCOPE

The data used in this paper were obtained from the 1950, 1960, and 1970 population censuses of the United States (U. S. Bureau of the Census, 1952, 1961, 1972). Except for minor adjustments in the figures from the three censuses,[1] these sources provide comparable data on occupational and industrial employment. The analyses are confined to data on the first-digit level of occupational and industrial changes in the employment of white and nonwhite residents in urban, rural nonfarm, and rural farm areas of Mississippi between 1950 and 1970.

Definitions or explanations of basic concepts or certain categories adopted in this study are those developed by the U. S. Bureau of the Census. These definitions are "substantially the same" in the three censuses, especially with respect to Mississippi, as indicated below.

Urban Residence

"The urban population consists of all persons living in (a) places of 2,500 inhabitants or more incorporated as cities, villages, boroughs (except Alaska), and towns (except in the New England States, New York, and Wisconsin), but excluding those persons living in the rural portions of extended cities; (b) unincorporated places of 2,500 inhabitants or more; and (c) other territory, incorporated or unincorporated, included in urbanized areas" (U. S. Bureau of the Census, 1972: App. 1-2).

Rural-Farm Residence

Rural-farm residence "comprises all rural residents living on farms of 10 or more acres from which sales of farm products amounted to $50 or more in the preceding calendar years or on places of less than 10 acres from which sales of farm products amounted to $250 or more in the preceding year." (U.S. Bureau of the Census, 1972: App.2).

Rural-Nonfarm Residence

Rural-nonfarm residents include all "persons in rural territory who did not meet the definition for the rural-farm population" (U. S. Bureau of the Census, 1972: App. 2).

Occupational Structure

Occupational structure refers to the allocation of manpower to various functions within a society contained in strata called occupational groups. As noted by Blau and Duncan (1967: 67), "the occupational structure in modern industrial society" shows "the allocation of manpower to various institutional spheres, and it is the flow of movements among occupational groups that reflects the adjustment of the demand for diverse services and supply of qualified manpower." Expressed as data, occupational structure is the number of persons employed according to the occupational classification system adopted by the U. S. Bureau of the Census. (El Attar and Saunders, 1974: 320. This paper treated the occupational classification system as composed of 11 occupational categories (Table 1).

Industrial Structure

Industrial structure refers to the distribution of a society's manpower among industries. Expressed as data, it is the number of persons employed according to industry (El Attar and Saunders, 1974: 32). The industrial classification system as used in this paper consisted of 15 major industry groups (Tables 2 and 3). The major industry group "professional services" included hospitals, health services, welfare, and religious and miscellaneous professional services.

Scope of the Analysis

It may be pointed out that the scope of the present analysis extended to only the civilian employed residents. It covered neither the armed forces nor the experienced unemployed residents.

THE ANALYSIS

The basic purpose of this paper is to investigate the extent to which occupational differentials in residential employment were the consequence of industrial differentiation. Specifically, the basic hypothesis is that in the process of social transformation and economic development, residential types tend to be differentiated with respect to the changes in the occupational structure of their employed residents. Underlying this hypothesis is the assumption that the occupational differences observed among the areas, are a product of differences in the changes in the industrial structure of each of these residential types.

Occupational Differentials

Table 1 shows that especially for the period between 1960 and 1970, urban and rural-nonfarm residential areas were attractive to employed residents of all occupations except farmers, private household workers, and nonwhite farm laborers and laborers. On the contrary, rural-farm areas were attractive to nonwhite residents employed in professional, managerial, and clerical occupations and to both white and nonwhite service workers in 1960 to 1970.

The pattern of change for 1950 to 1960 was entirely different from that for 1960 to 1970, except for rural-farm areas and, to some extent, for nonwhite residents. While the urban and rural-nonfarm residential areas showed gains in white - and non-white employed dwellers, rural-farm areas incurred losses. Differentiation within and between white and nonwhite employed residents in this regard was very apparent during the period under study. The gain for whites and nonwhites in urban and rural-nonfarm areas occurred at a decreasing rate; rural farm losses occurred at a constant rate for whites and at an increasing rate for nonwhites (see Table 1).

In order to ascertain whether the numerical differentials in occupational change of white and nonwhite employment observed during the period under study among the three residential types were statistically significant or not, a three-way classification of analysis of variance was applied to the data in Table 1. This model was used to test the main effects of occupation (eleven levels), type of residence (three levels), and color (two levels). The data consisted of two observations (percent of occupational change between 1950 to 1960 and 1960 to 1970) on each of 66 different treatment combinations (occupation, type of residence, and color). The design was assumed to fit the following statistical model:

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk} ,$$

where

i = occupational group levels = 1, 2, . . ., 11; j = type of residence levels = 1, 2, 3; k = color levels = 1, 2; and l = time levels = 1, 2 (Dunn and Clark, 1974: 145-154).

Table 4 summarizes analysis of variance for the data on occupational change in Table 1, namely the F tests of seven null hypotheses (H_n's). These hypotheses were stated as follows: For main effects, three H_n's were tested: H_n 1: effects due occupation change = 0 ($\Sigma\alpha_i = 0$); H_n 2 effects due residence = 0 ($\Sigma\beta_j = 0$); effects due color = 0 ($\Sigma\gamma_k = 0$); H_n 4, H_n 5 and H_n 6 which refer to effects due two-factor interaction = 0 each; and H_n 7 effects due three-factor interaction = 0. Based on the calculated variance ratios shown in the last column of Table 4, H_n 1, H_n 2 and H_n 6 were rejected; the first two at the .01 level and the last at .005 level. Accordingly, one might conclude that the observed differences among the three residential type areas with regard to occupational changes were statistically significant. On the other hand, the differences between white and nonwhite residents with respect to occupational employment changes were not significant. However, type of residence and color were not independent. Finally, one might conclude that the effect of occupation, type of residence, and color (three-factor interaction) were independent.

Industrial Differentials

Tables 2 and 3 provide data on the number of and percent changes in the white and nonwhite industrial employment of Mississippi residents 14 years old and over in 1950, 1960, and 1970 by types of residence. The total percent changes in the number of employed persons in 1950-60 and in 1960-70 were -4.8 and 6.3, respectively. The loss which occurred in 1950-60 was caused solely by negative change in the nonwhite employment which amounted to -20.3 percent in comparison with an increase of 7.4 percent in the size of white employment. On the contrary, employment growth in 1960-70 was mainly due to an increase in the white employment by 20.2 percent as compared to -17.3 percent for the nonwhites. Of all three residential types, only the rural - farm areas showed negative change in the size of employment amounting to -48.1 and -47.7 for the whites and -59.7 and -74.6 for the nonwhites in 1950-60 and 1960-70, respectively. In the 15 major industry groups, especially in 1960-70, rural farm areas had negative employment changes for whites and nonwhites (Table 3), with exception of finance, entertainment (whites), educational services (nonwhites), professional services, and public adminstration (nonwhites).

In order to assess the extent to which occupational differentials in residential employment were promoted by industrial differentiation, a three-way variable model of analysis of variance, similar to the one used above, was applied to the data on changes in industrial employment in Table 3. Table 5 provides a summary for the F tests of seven H_n 's. Only three H_n 's (variations due industries, residence, and residence and color) were significant at .001 level. It is of interest to indicate that in the two analyses of variance (occupation and industry) residence and color were significantly dependent. However, this dependence disappeared once the effect of industry was included in the analysis, implying the tendency of industrial differentiation to promote racial integration.

CONCLUSION

This paper fulfilled three objectives: First, it examined and compared residential differentials in occupational changes of white and nonwhite employment; second, it assessed the statistical significance of these differentials; and third it asserted the assumption that the differentiations in the occupational structure of the employed residents of the three residential types were a product of the differences in the industrial structure of each residential type. Put differently, the residential pattern of employed Mississippians was a function of the changes in the occupational structure which followed the changes in the pattern of industrial differentiation. On the other hand, one should not neglect the effect of commuting, annexation, and migration on the process of differentiation in people and jobs in the state.

Finally, the analysis suggests that Mississippi should not be viewed as a rural state whose urban environment does not differ from the rural aggregates in terms of the occupational composition of its employed residents. Such a view is spurious and does not represent the existing reality.

FOOTNOTES

*The research on which this paper was based was part of Mississippi Agricultural and Forestry Experiment Station Population Project No. 4004. The author gratefully acknowledges the assistance of Mr. David L. Steinman.

[1] The comparability of the 1970 data on employed persons classified by occupations and industries is affected by three types of changes in the 1970 census procedures, namely: (1) the age coverage of employed persons is limited to persons 16 years old and over, (2) title and content of certain occupations/industries, and (3) allocation of the "not reported" cases. The 1970 data were adjusted with regard to the first two items, for example, the 14 and 15 years old employed persons were proportionately distributed among the occupations/ industries by sex and color, and the occupation category "transport equipment operatives" was included in "operatives." Another adjustment was related to discrepancies effected in the 1970 "rural-nonfarm" and "rural-farm" categories as corrected by the U.S. Bureau of the Census. The correction resulted in making some occupational group for each color did not compare with the total as originally given in the census. The Population Division of the U. S. Bureau of the Census (Mr. Speaker) advised the author to adjust for such discrepancies by relating all adjustments to the "rural-nonfarm" category. The "not reported" cases in 1950 and 1960 data were proportionately distributed among the occupations/ industries. This category numbered, 1950 followed by 1960 figures, 5,792 and 10,968 for white occupational employment; 4,028 and 6,119 for nonwhite occupational employment; 6,323 and 8,259 for white industrial employment; and 4,472 and 5,198 for nonwhite industrial employment. According to type of residence, the figures were as follows:

	Urban		Rural Nonfarm		Rural Farm	
Year	White	Non White	White	Non White	White	Non White
Occupation						
1950	1,551	873	1,736	833	2,505	2,322
1960	5,627	3,182	3,608	1,727	1,733	1,210
Industry						
1950	1,691	950	1,882	989	2,750	2,533
1960	4,167	2,564	2,611	1,542	1,481	1,092

The size of employed persons 14 and 15 years old in 1970 amounted to 4,043 for the whites and 1,706 for the nonwhites. Distribution by type of residence (taken in the followin order: urban, rural nonfarm, and rural farm) was: 2,104, 1,649, and 290 for white persons; and 777, 764, and 165 for nonwhite persons.

TABLE 1. EMPLOYED POPULATION 14 YEARS OLD AND OVER BY MAJOR OCCUPATION GROUP, TYPE OF RESIDENCE AND COLOR IN MISSISSIPPI: 1950, 1960, AND 1970.

Occupation and Color		Number of Employed									Percent Change in Employment					
		Urban			Rural Nonfarm			Rural Farm			Urban		Rural Nonfarm		Rural Farm	
		1950	1960	1970	1950	1960	1970	1950	1960	1970	50-60	60-70	50-60	60-70	50-60	60-70
Professional, etc.	(W	18,746	27,862	45,591	10,968	14,138	21,589	4,397	3,666	3,265	48.63	63.63	28.90	52.70	-16.61	-10.94
	(NW	3,864	6,236	10,422	2,048	2,955	6,490	1,988	1,239	1,342	61.39	67.13	44.29	119.63	-37.68	8.31
Farmers, etc.	(W	1,117	1,186	1,004	3,816	7,241	4,412	91,853	31,798	10,780	6.18	-15.35	89.75	-39.07	65.38	-66.10
	(NW	782	489	162	2,905	7,481	1,516	108,684	25,449	3,058	-37.47	-66.87	157.52	-79.74	-76.58	-87.98
Managers, etc.	(W	22,980	30,711	32,820	13,239	15,526	18,438	4,442	3,378	2,667	3.64	6.87	17.27	18.76	-23.95	-21.05
	(NW	2,149	1,823	2,307	967	983	1,339	243	159	201	-15.17	26.55	1.66	36.22	-34.57	26.42
Clerical wkrs.	(W	24,540	36,234	51,269	8,986	15,485	26,928	3,824	4,418	4,192	47.65	41.49	72.32	73.90	15.53	-5.12
	(NW	1,219	1,687	5,551	382	599	2,625	144	133	307	38.39	229.05	56.81	338.23	-7.64	130.83
Sales wkrs.	(W	18,224	21,563	25,628	9,375	11,185	12,595	3,945	3,252	1,849	18.32	18.85	19.31	12.61	17.57	-43.14
	(NW	1,173	1,148	1,687	604	543	678	209	160	139	-2.13	46.95	-10.10	24.86	-23.44	-13.12
Craftsmen	(W	20,245	26,484	34,632	16,139	26,333	41,523	7,849	7,375	5,526	30.82	30.77	63.16	57.68	-6.04	-25.07
	(NW	6,438	6,973	9,039	2,501	3,844	7,525	1,127	1,387	1,027	8.31	29.63	53.70	95.76	23.07	-25.96
Operatives	(W	20,869	29,431	34,665	19,315	41,925	61,669	14,273	17,338	10,142	41.03	17.78	117.06	47.09	21.47	-41.50
	(NW	17,582	19,313	23,672	8,730	11,546	23,005	5,615	3,801	3,447	9.84	22.57	32.26	99.25	-32.31	- 9.31
Pvt. hsehld. wkrs.	(W	490	1,399	1,079	436	1,462	1,114	236	576	154	185.51	-22.87	235.32	-28.80	144.07	-73.26
	(NW	20,214	25,833	15,231	8,779	14,719	10,928	3,510	5,253	1,434	27.80	-41.04	67.66	-25.76	49.66	-72.70
Serv. wkrs.	(W	9,355	12,822	20,538	4,954	8,319	15,052	1,892	1,857	2,318	37.06	60.18	67.92	80.94	- 1.85	24.82
	(NW	14,747	18,685	20,448	4,436	6,289	11,516	1,194	1,350	1,476	26.70	9.44	41.77	83.11	13.07	9.33
Farm laborers	(W	349	354	733	3,094	4,545	5,071	24,601	7,789	2,150	1.43	107.06	46.90	11.57	-68.34	-72.40
	(NW	1,518	1,691	1,340	8,162	21,317	12,087	52,603	30,267	4,530	11.40	-20.76	161.17	-43.30	-42.46	-85.03
Laborers	(W	3,953	4,340	6,800	7,015	7,477	9,966	5,071	2,805	1,430	9.79	56.68	6.59	33.29	-44.69	-49.02
	(NW	16,090	14,014	10,665	10,645	11,201	11,697	5,011	3,498	1,512	-12.90	-23.90	5.22	4.43	-30.19	-56.78
Total	(W	140,868	192,386	254,759	97,337	153,636	218,357	162,383	84,252	44,073	36.57	32.42	57.84	42.13	-48.12	-47.69
	(NW	85,776	97,892	100,524	50,159	81,477	89,406	180,328	72,696	18,473	14.13	2.69	62.44	9.73	-59.69	-74.59

Source: 1950 figures are from U. S. Bureau of the Census, U. S. Census of Population: 1950, General Characteristics, Mississippi (Washington, D. C.: Government Printing Office, 1952), Table 28A, p. 34; 1960 figures are from U. S. Census of Population: 1960, General Social and Economic Characteristics, Mississippi (1961), Table 58, P. 126; 1970 figures adjusted to 1960 age coverage and for discrepancies in the figures for "rural farm" are from U. S. Census of Population: 1970, General Social and Economic Characteristics, Mississippi (1972), Table 54, pp. 160-161 and the corresponding tables on "rural nonfarm" and "rural farm" data as corrected by the U.S. Bureau of the Census. Mr. Robert C. Speaker, Population Division, U.S. Bureau of the Census has advised the author to relate any discrepancies in the data published in these tables to the "rural nonfarm" category. The 1950 and 1960 figures were adjusted for "not reported."

TABLE 2. EMPLOYED PERSONS 14 YEARS OLD AND OVER BY MAJOR INDUSTRY GROUP, TYPE OF RESIDENCE, AND COLOR IN MISSISSIPPI: 1950, 1960, AND 1970

	Urban						Rural Nonfarm						Rural Farm					
	White			Nonwhite			White			Nonwhite			White			Nonwhite		
Industry	1950	1960	1970	1950	1960	1970	1950	1960	1970	1950	1960	1970	1950	1960	1970	1950	1960	1970
Agric.	2,909	3,053	3,542	3,166	3,227	2,232	8,763	14,603	12,154	12,296	30,247	14,861	118,523	41,075	13,648	164,190	56,615	7,880
Mining	1,564	2,539	2,585	131	177	217	1,330	2,660	3,969	119	121	322	489	523	407	41	66	61
Constr.	11,090	12,948	16,471	6,521	6,778	6,875	9,731	15,234	21,181	2,172	3,621	5,615	6,462	5,675	3,395	1,039	1,502	814
Manuf.	23,351	39,560	50,902	14,546	14,600	21,870	21,474	45,468	74,370	11,358	12,744	25,450	15,007	17,283	11,353	6,020	3,763	3,543
Transp.	13,792	15,734	17,983	5,123	4,626	5,025	6,029	8,836	13,424	2,459	2,755	4,283	2,876	2,915	1,863	936	1,120	748
Trade	42,160	50,171	61,448	15,385	16,859	14,271	24,019	30,319	40,342	5,236	7,000	7,862	8,884	7,456	5,173	1,320	1,455	948
Fin.	6,018	10,987	15,756	799	963	1,241	1,653	2,834	5,139	179	195	342	573	631	788	32	28	32
Bus. serv.	1,265	1,749	3,172	98	194	443	276	572	1,195	11	38	145	87	119	99	5	16	14
Repr. serv.	2,985	3,271	3,866	1,219	1,049	1,071	3,173	3,203	4,014	661	608	745	1,259	709	367	266	141	95
Pvt. hsld.	868	1,757	1,158	22,853	27,782	14,558	717	1,771	1,247	9,842	15,772	10,925	351	676	181	3,710	5,522	1,418
Pers. serv.	6,310	7,712	8,687	7,321	7,527	7,513	2,962	3,771	5,033	1,719	1,800	2,634	856	744	592	365	407	245
Entrmnt.	1,788	1,396	1,401	712	760	578	865	689	660	156	173	171	154	74	83	27	45	24
Ed. serv.	7,187	11,756	23,555	4,043	7,540	14,035	7,464	8,895	12,926	2,488	4,258	10,390	3,442	2,753	2,346	1,972	1,550	1,923
Prof. serv.	9,959	16,212	27,498	2,922	4,415	8,346	3,978	7,388	14,365	1,166	1,746	4,601	1,018	1,371	2,129	292	363	611
Pub. admin.	9,622	13,541	16,735	937	1,395	2,249	4,903	7,393	8,338	297	399	1,060	2,402	2,248	1,649	113	103	117
Total	140,868	192,386	254,759	85,776	97,892	100,524	97,337	153,636	218,357	50,159	81,477	89,406	162,383	84,252	44,073	180,328	72,696	18,473

Source: Same census reports cited in Table 1: 1950 figures are from Table 30a, pp. 36-37; 1960 figures are from Table 61, pp. 129-130; and 1970 figures are from Table 55, pp. 162-163. The 1950 and 1960 figures were adjusted for "not reported"; 1970 figures were adjusted to include employed population 14 and 15 years old.

TABLE 3. PERCENT CHANGE IN EMPLOYED PERSONS 14 YEARS OLD AND OVER BY MAJOR INDUSTRY GROUP, TYPE OF RESIDENCE, AND COLOR IN MISSISSIPPI: 1950-1970

	Urban				Rural Nonfarm				Rural Farm			
	White		Nonwhite		White		Nonwhite		White		Nonwhite	
Industry	50-60	60-70	50-60	60-70	50-60	60-70	50-60	60-70	50-60	60-70	50-60	60-70
Agric., etc.	4.95	16.02	1.93	-30.83	66.64	-16.77	145.99	-50.87	-65.34	-66.77	-65.52	-86.08
Mining	62.34	1.81	35.11	22.60	100.00	49.21	1.68	166.12	6.82	-22.18	60.98	-7.58
Constr.	16.75	27.21	3.94	1.43	56.55	39.04	66.71	55.07	-12.18	-40.18	44.56	-45.81
Manuf.	69.41	28.67	0.37	49.79	111.74	63.57	12.20	99.70	15.17	-34.31	-37.49	-5.85
Transp., etc.	14.08	14.29	9.70	8.63	46.56	51.92	12.04	55.46	1.36	-36.09	19.66	-33.21
Trade	19.00	22.48	9.58	-15.35	26.23	33.06	33.69	12.31	-16.07	-30.62	10.23	-34.85
Fin.	82.57	43.41	20.53	28.87	71.45	81.33	8.94	75.38	10.12	24.88	-12.50	14.29
Bus. serv.	38.26	81.36	97.96	128.35	107.25	108.92	245.45	281.58	36.78	-16.81	220.00	-12.50
Repr. serv.	9.58	18.19	-13.95	2.10	0.95	25.32	-8.02	22.53	-43.69	-48.24	-46.99	-32.62
Pvt. hsld.	102.42	-34.09	21.57	-47.60	147.00	-29.59	60.25	-30.73	92.59	-73.22	48.84	-74.32
Pers. serv.	22.22	12.64	2.81	-0.19	27.31	33.47	4.71	46.33	-13.08	-20.43	11.51	-39.80
Entrmnt.	-21.82	0.36	6.74	-23.95	-20.35	-4.21	10.90	-1.16	-51.95	12.16	66.67	-46.67
Ed. serv.	63.57	100.37	86.50	86.14	19.17	45.32	71.14	144.01	-20.02	-14.78	-21.40	24.06
Prof. serv.	62.79	69.62	51.10	89.04	85.72	94.44	49.74	163.52	34.68	55.29	24.32	68.32
Pub. admin.	40.73	23.59	48.88	61.22	50.79	12.78	34.34	165.66	-6.41	-26.65	-8.50	13.59
Total	36.57	32.42	14.13	2.69	57.84	42.13	62.44	9.73	-48.12	-47.69	-59.69	-74.59

Source: Computed from Table 2 above.

TABLE 4. ANOVA FOR CHANGES IN THE OCCUPATIONAL STRUCTURE OF EMPLOYED URBAN, RURAL NONFARM, AND RURAL FARM RESIDENTS IN MISSISSIPPI, BY COLOR: 1950-1970

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	Computed F
(A) Due occupation	95,040.7412	10	9,504.0741	2.698[a]
(B) Due residence	36,869.8193	2	18,434.9096	5.234[a]
(C) Due color	3,760.8452	1	3,760.8452	1.068[b]
Due AB	34,756.8164	20	1,737.8408	0.493[b]
Due AC	7,327.8701	10	732.7870	0.208[b]
Due BC	69,242.0635	2	34,621.0318	9.829[c]
Due ABC	x	x	x	x
Residual	302,933.4735	86	3,522.4822	
Total	549,931.6292	131		

[x]Sum of squares due ABC (41,411.7383, with 20 degrees of freedom) were added to the residuals (261,521.7352) because of their insignificant contribution (F = .523).

[a]Significant at .01 level. [b]Not significant. [c]Significant at .005 level.

TABLE 5. ANOVA FOR CHANGES IN THE INDUSTRIAL STRUCTURE OF EMPLOYED URBAN, RURAL NONFARM, AND RURAL FARM RESIDENTS IN MISSISSIPPI, BY COLOR: 1950-1970

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	Calculated F
(A) Due industries	177,263.1484	14	12,661.6535	6.472[a]
(B) Due residence	40,754.7812	2	20,377.3906	10.417[a]
(C) Due color	700.1389	1	700.1389	0.358[b]
Due AB	47,112.3750	28	1,682.5848	0.860[b]
Due AC	2,422.1152	14	173.0082	0.088[b]
Due BC	86,314.5840	2	43,157.2920	22.061[a]
Due ABC	x	x	x	x
Residual	230,835.0856	118	1,956.2295	
Total	585,402.2283	179		

[x]Sum of squares due ABC (46,713.7852, with 28 degrees of freedom) were added to the residuals (194,121.3004) because of their insignificant contribution (F = .816).

[a]Significant at .001. [b]Not significant.

REFERENCES

Blau, Peter M. and Otis D. Duncan. 1967. The American Occupational Structure. New York: John Wiley.

El Attar, M. E. 1974. Some Aspects of Population Dynamics in Mississippi. MAFES Bulletin 814. Mississippi State: Mississippi Agriculture and Forestry Experiment Station.

El Attar, M. E. and J. Saunders. 1974. "Mississippi Jobs, Measuring Components of Occupational Change." Growth and Change. 5(October), 32-38.

U. S. Bureau of the Census. 1952. U. S. Census of Population: 1950. Vol. II. Characteristics of Population. Mississippi. Washington, D. C.: U. S. Government Printing Office.

U. S. Bureau of the Census. 1961. U. S. Census of Population: 1960. General Social and Economic Characteristics. Mississippi Final Report PC (1)-26C. Washington, D.C.: U. S. Government Printing Office.

U. S. Bureau of the Census. 1972. U. S. Census of Population: 1970. General Social and Economic Characteristics. Final Report PC(1)-C26 Mississippi. Washington, D.C.: U. S. Government Printing Office.

ON LOG-LINEAR MODELS FOR MULTIPLE-RECORD SYSTEMS

M. Nabil El-Khorazaty, Gary G. Koch and Alcinda L. Lewis
University of North Carolina at Chapel Hill
Peter B. Imrey, University of Illinois

1. Introduction

Estimating the size of a population of humans, animals, or events relating to a community of such units is a major methodological problem shared by researchers in many disciplines. Where counts of organisms are involved, we may be concerned with such quantities as the total number of residents, victims of a congenital anomaly, criminals or victims of crimes, drug abusers, parties involved in automotive accidents, or of fish in a lake, deer in a forest, or bacteria on a microscope slide. Populations of events frequently studied are those of births, deaths, migrations, marriages, separations, divorces, and diagnoses of cancer or other diseases within a chosen time interval. Frequently, the ultimate aim is not only estimation of magnitude, as shown by a total, but rather of change over time as shown by a rate of population growth or the incidence rate of an event such as disease attack.

In principle, most data required to achieve reasonable aims for the study of human populations can be supplied by census (which covers the whole population) and civil registration systems (continuous recording of events of interest), supplemented by periodic sample surveys during intercensal periods. However, these traditional statistical systems are not adequate in many countries of the world and, for obvious reasons, are unsuitable for use with animal populations. Inadequacies of the traditional sources have led to the development of statistical procedures designed to combine the information available from multiple sources or protocols for detecting the same individuals or events, when each protocol is known to be insufficiently sensitive if used by itself. For human populations these procedures, generally grouped under the term Multiple-Record Systems (MRS), attempt to estimate the number of persons or events while adjusting for the individual fallibilities of census, survey or vital registration systems. Analogous "Capture-Mark-Recapture" (CMR) techniques attempt to compensate for the virtual impossibility of locating and distinguishing, in a short time, even the majority of members of any interesting animal population. In practice, CMR is distinguished by the use of i) detection protocols which are not coterminous in time, and ii) labeling and subsequent identification procedures which are unavailable for human studies.

The Multiple-Record System (MRS) involves data collection from two or more sources of information (called recording systems) which cover the same sample or sub-sample of areas and the same time period. The special case of two sources (Dual-Record System (DRS)) has been used widely in the last 30 years to adjust for omissions in the recording of vital events and to estimate population growth rates. In this regard, Chandrasekar and Deming (1949) present a theoretical framework for estimating the total number of events under the following Assumptions (1)-(3).

1. No coverage errors with respect to the scope of area and/or time period in which individuals or events are recorded (i.e., each system only records individuals or events that pertain to the target area and/or time period under study);
2. Independence of recording systems (i.e., the probability that both systems detect a randomly chosen individual or event is the product of the probabilities of detecting a randomly chosen element for each system individually);
3. No misclassification errors with respect to determining exactly which systems have detected an individual or event (i.e., a perfect matching rule exists for linking information from the two systems to determine correctly the number of individuals or events detected by both).

Chakraborty (1963) and Das Gupta (1964) extend this approach to situations involving three or more sources of information.

A fruitful approach to the study of MRS and CMR data is to view the involved set of detection protocols (recording systems for MRS, capture or observation methods or times for CMR) as a probabilistic process, or channel, with an input and resulting output. Input to the processor is a single element of the population to be studied, while the resulting output is a response pattern which delineates exactly which of the various detection protocols (if any) have recorded, or captured, that element. The aggregate result of passing every population element through the processor may be arranged, for d detection protocols, in a 2^d contingency table with each dimension, or marginal, describing the success or failure of a single protocol in capturing the elements of the target population. In a typical MRS or CMR application, we see this contingency table absent the single cell containing those elements missed by all detection protocols. It is this "incomplete contingency table" which must be used to generate estimates of population size. Sometimes the population under study may be partitioned into a set of subpopulations according to such demographic variables as geographic location, urbanization, or sex, so that we see such an incomplete table with its missing cell for each subpopulation.

Fienberg (1972), Bishop, Fienberg and Holland (1975), El-Khorazaty (1975) and Koch, El-Khorazaty and Lewis (1977) advocate fitting log-linear models to the observed cells of the above tables, and using these models to obtain refined estimates of the missing cell(s) and, hence, of the population total. Such refined estimates may be obtained by

i) controlling for statistical dependence of specific types among the actions of the various detection protocols;
ii) accounting for or modeling the effects of subpopulations or their defining factors

on the probabilistic properties of the detection protocols.

This generalization of the Chandrasekar-Deming approach for the DRS and the Peterson-Lincoln approach for the CMR allows great latitude in choosing an estimation procedure realistically adapted to the properties of actual recording systems.

In this paper we give a matrix formulation of the general log-linear model applicable to data obtained from the operation of a multiple-record system on a stratified population. The matrix formulation yields explicit matrix product expressions for the true and estimated asymptotic covariance matrices of efficient estimators for the log-linear model parameter vector, as well as corresponding results for the asymptotic covariances of fitted detection probabilities in the several strata, and of the stratum-specific inflation factors used to estimate the stratum sizes.

2. Matrix Formulation of the Log-Linear Model

2.1. Notation

Let $i = 1,2,\ldots,s$ index a set of sub-populations (or strata). Let $g = 1,2,\ldots,d$ index a set of recording systems, and $j_g = 1,2$ represent the presence $(j_g = 1)$ or absence $(j_g = 2)$ of the attribute corresponding to registration by the g-th recording system. Let the vector subscript $\underset{\sim}{j}' = (j_1,j_2,\ldots,j_d)$ index the multivariate response profiles for simultaneous recording status with respect to the d recording systems. Let $n_{i;j_1j_2\ldots j_d}$ denote the number of elements from the i-th sub-population with recording status $(j_1,j_2,\ldots,j_d)$. Let

$$p_{i;j_1j_2\ldots j_d} = n_{i;j_1j_2\ldots j_d}/n_i \qquad (2.1.1)$$

where n_i is the total number of elements of sub-population $i\ (i = 1,2,\ldots,s)$ recorded by any of the record systems. Let

$$\bar{\pi}_{i;j_1j_2\ldots j_d} = \frac{\pi_{i;j_1j_2\ldots j_d}}{1 - \pi_{i;22\ldots 2}} \qquad (2.1.2)$$

where $\pi_{i;j_1j_2\ldots j_d}$ is the probability that a population element has recording status $\underset{\sim}{j}$ for sub-population i. Thus, $\bar{\pi}_{i;j_1j_2\ldots j_d}$ denotes the conditional probability that a population element has recording status $\underset{\sim}{j}$ for sub-population i, given that it is observed.

2.2. General Procedure for Estimating Population Size

(a) Formulate a rank $t \le s(2^d-2)$ log-linear model for the $\pi_{i;j_1j_2\ldots j_d}$;

(b) Estimate the parameters of the model from the unrestricted maximum likelihood estimators $p_{i;j_1j_2\ldots j_d}$ of $\bar{\pi}_{i;j_1j_2\ldots j_d}$ provided by the observed incomplete contingency table;

(c) Obtain estimates $\hat{\pi}_{i;j_1j_2\ldots j_d}$ of $\pi_{i;j_1j_2\ldots j_d}$ from the fitted model;

(d) Estimate N_i, the size of stratum i, by $\dfrac{n_i}{1-\hat{\pi}_{i;22\ldots 2}}$.

When $s = 1$, $d = 2$ and the model chosen in (a) contains only main effects corresponding to the two record systems (which are thus assumed to operate independently in the sense of no association of detection by the two systems), the above procedure yields the classical Chandrasekar-Deming estimate. When $s = 1$ and $d > 2$, the choice of a model with no interaction terms, but with main effects corresponding to each system, yields the extension of their estimator derived by Chakraborty (1963).

2.3. Representation of the Model

For any specific stratum i, a general log-linear model for the corresponding $\pi_{i;j_1j_2\ldots j_d}$ may be written as

$$\underset{\sim}{\pi}_i = \underset{\sim}{\pi}_i(\underset{\sim}{\beta}) = \underset{\sim}{\exp}(\underset{\sim}{X}_i\underset{\sim}{\beta})/\underset{\sim}{1}_r'\underset{\sim}{\exp}(\underset{\sim}{X}_i\underset{\sim}{\beta}) \qquad (2.3.1)$$

where $\underset{\sim}{\pi}_i$ is the vector of the $\pi_{i;\, j_1j_2\ldots j_d}$ arranged in lexicographic order, $\underset{\sim}{X}_i$ is a known "design matrix" specifying the structure of the log-linear model, $\underset{\sim}{\beta}$ is the (unknown) vector of $t \le (r-2)$ model parameters, $\underset{\sim}{1}_k$ is a k-vector of units, $\underset{\sim}{\exp}$ is the elementwise exponential operator, and $r = 2^d$. $\underset{\sim}{X}_i$ is assumed to be of full rank, with columns jointly linearly independent of the vector $\underset{\sim}{1}_r$ representing the underlying linear restriction that $\sum_{\underset{\sim}{j}}\pi_{i;\underset{\sim}{j}} = 1$. Fienberg (1972) and Bishop, Fienberg and Holland (1975) further restrict to the class of hierarchical analysis of variance models, due to the ease of obtaining the maximum likelihood estimate $\tilde{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$, under the conditional multinomial likelihood for the $n_{i\underset{\sim}{j}}$, through the computational technique of "iterative proportional fitting". These models correspond to design matrices $\underset{\sim}{X}_i$ for which the set of columns of $\underset{\sim}{X}_i$, $[\underset{\sim}{X}_i]_c$, can be written as

$$[\underset{\sim}{X}_i]_c = \bigcup_{k=1}^{K}[\underset{\sim}{X}_{ik}]_c \quad \text{for some } K ,$$

where each $\underset{\sim}{X}_{ik}$ is the usual design matrix corresponding to a complete (or saturated) factorial model involving as factors some subset of the d record systems. Results of this section apply to general $\underset{\sim}{X}_i$; we adopt the conditional multinomial likelihood, but address directly the problem of obtaining estimates only in Section 4.

For $s > 1$, we may generalize the above formulation to a model for $\underset{\sim}{\pi}' = (\underset{\sim}{\pi}_1',\underset{\sim}{\pi}_2',\ldots,\underset{\sim}{\pi}_s')$ as

$$\underset{\sim}{\pi} = \underset{\sim}{\pi}(\underset{\sim}{\beta}) = \underset{\sim}{D}_{\underset{\sim}{\eta}}^{-1}\{\underset{\sim}{\exp}\,\underset{\sim}{X}\,\underset{\sim}{\beta}\} . \qquad (2.3.2)$$

Here the vector $\underset{\sim}{\beta}$ of unknown model parameters is of dimension $t \le s(r-2)$, and underlies the joint detection probabilities for all strata through the composite design matrix

$$\underset{rs\times t}{\underset{\sim}{X}'} = (\underset{\sim}{X}_1',\underset{\sim}{X}_2',\ldots,\underset{\sim}{X}_s') .$$

For general $\underset{\sim}{y}$, $\underset{\sim}{D}_{\underset{\sim}{y}}$ is the diagonal matrix with $\underset{\sim}{y}$ on the principal diagonal, and

$$\underset{\sim}{\eta} = [\underset{\sim}{1}_{rr} \otimes \underset{\sim}{I}_s]\ \underset{\sim}{\exp}(\underset{\sim}{X}\underset{\sim}{\beta})$$

with $\underset{\sim}{1}_{vw}$, $\underset{\sim}{I}_v$ and $\otimes$ representing a $v \times w$ matrix of units, the $v \times v$ identity and Kronecker multiplication respectively. Each of the $\underset{\sim}{X}_i$ is assumed of full rank, with columns jointly linearly independent of $\underset{\sim}{1}_r$. Otherwise the $\underset{\sim}{X}_i$ may be of essentially free form and vary considerably from stratum to stratum. In particular, some columns of X_i may be 0, indicating that certain parameters of β apply only to certain strata. Clearly, a model with small t is desirable if realistic.

3. Determination of Covariance Structure

As noted previously, we assume for the generated cell counts a multinomial distribution conditional on the totals n_i of elements detected in stratum i by any of the record systems. Thus, the joint likelihood ϕ may be written as

$$\phi = \prod_{i=1}^{s}\left[n_i! \prod_{j\neq(2,2,\ldots,2)} \pi_{i;j}^{n_{i;j}}/n_{i;j}!\right], \quad (3.1)$$

with the s constraints that

$$\sum_{j\neq(2,2,\ldots,2)} \pi_{i;j} = 1 \quad \text{for all } i = 1,2,\ldots,s \,.$$

Incorporating the log-linear model to this likelihood and expressing the result in matrix terms yields

$$\phi = \frac{\prod_{i=1}^{s} n_i!}{\prod_{i=1}^{s}\prod_{j\neq(2,2,\ldots,2)} n_{i;j}!} \; \frac{\exp(n_0' X_0 \beta)}{\prod_{i=1}^{s}\{1_{(r-1)}' \exp(X_{i0}\beta)\}^{n_i}} \quad (3.2)$$

where $n_0' = (n_{10}', n_{20}', \ldots, n_{s0}')$, n_{i0} is the vector of observed cell counts from the i-th stratum incomplete table (lexicographic order), X_{i0} is the matrix consisting of the first $(r-1)$ rows of X_i, and

$$X_0' = (X_{10}', X_{20}', \ldots, X_{s0}').$$

The asymptotic covariance matrix of any asymptotically efficient (such as maximum likelihood or minimum Neyman chi-square) estimate $\hat{\beta}$ of β is available as the negative inverse of Fisher's Information Matrix. Using matrix differentiation methods similar to those of Forthofer and Koch (1973), we obtain

$$\frac{d}{d\beta}[\log \phi] = -n_0' X - [1_{(r-1)} \otimes n_.]' D_{\Pi_0(\beta)} X_0 \,, \quad (3.3)$$

where $n_.' = (n_1, n_2, \ldots, n_s)$ and $\Pi_0(\beta)$ is defined analogously to n_0 from the $\pi_{i;j}$. Hence, the asymptotic covariance $V_{\hat{\beta}}(\Pi_0)$ of $\hat{\beta}$ is obtained as

$$V_{\hat{\beta}}(\Pi_0) = \left\{\frac{-d^2}{d\beta d\beta'}[\log \phi]\right\}^{-1}$$

$$= \left\{\sum_{i=1}^{s} n_i X_{i0}'[D_{\Pi_0} - \Pi_0 \Pi_0'] X_{i0}\right\}^{-1}, \quad (3.4)$$

following simplifications. Since $\hat{\Pi}_0 = \Pi_0(\hat{\beta})$ is consistent for $\Pi_0 = \Pi_0(\beta)$, a consistent estimator for the covariance matrix $V_\beta(\Pi_0)$ is

$$V_{\hat{\beta}} = V_{\hat{\beta}}(\hat{\Pi}_0) = \left\{\sum_{i=1}^{s} n_i X_{i0}'[D_{\hat{\Pi}_0} - \hat{\Pi}_0 \hat{\Pi}_0'] X_{i0}\right\}^{-1}. \quad (3.5)$$

The asymptotic covariance matrix for the estimator $\hat{\Pi}_0$ of the vector of conditional probabilities of the response profiles for the various strata, and the estimators $\hat{\gamma}_i = \gamma_i(\hat{\beta})$ of the stratum-specific ratios

$$\gamma_i = \frac{\pi_{i;22\ldots2}}{1-\pi_{i;22\ldots2}} = \frac{\pi_{i;22\ldots2}}{[1_{(r-1)}', 0]\Pi_i} \,,$$

is obtained by use of the well-known δ-method as based on the first-order Taylor series approximations for these estimators. In this regard, the compound function notation used in Forthofer and Koch (1973) is used to express $\hat{\Pi}_0$ and the $\hat{\gamma}_i$ in the form

$$\hat{\theta}' = (\hat{\Pi}_{10}, \hat{\gamma}_1, \hat{\Pi}_{20}, \hat{\gamma}_2, \ldots, \hat{\Pi}_{s0}, \hat{\gamma}_s)$$

as $\hat{\theta} = \theta(\hat{\beta}) = \exp[A_3 \log\{A_2 \exp(A_1 \hat{\beta})\}]$ where $A_1 = X_0$,

$$A_2 = \begin{bmatrix} I_r \\ 1_{(r-1)}', \; 0 \end{bmatrix} \otimes I_s \,,$$

$$A_3 = [I_r, \; -1_r] \otimes I_s \,,$$

and $\log$ is the elementwise logarithmic operator. As a result, a consistent estimator for the corresponding asymptotic covariance matrix can be determined as the matrix product

$$\hat{V}_{\hat{\theta}} = D_{\chi_3} A_3 D_{a_2}^{-1} A_2 D_{\chi_1} A_1 [V_{\hat{\beta}}(\hat{\Pi}_0)] A_1' D_{\chi_1} A_2' D_{a_2}^{-1} A_3' D_{\chi_3} \quad (3.6)$$

where $\chi_1 = \exp(A_1\hat{\beta})$, $a_2 = A_2\chi_1$, $\chi_3 = \exp\{A_3 \log(a_2)\}$.

The approach ultimately leads as described in Section 2.2, to the estimators

$$\hat{n}_{i;22\ldots2} = n_i \hat{\gamma}_i$$

and

$$\hat{N}_i = n_i(1 + \hat{\gamma}_i)$$

for the missing cell and total size of stratum i, and

$$\hat{N} = \sum_{i=1}^{s} n_i(1 + \hat{\gamma}_i)$$

for the population size. Since the n_i are random variables assumed to have independent binomial distributions with parameters N_i and $(1-\pi_{i;22\ldots2})$, the methods indicated in Darroch (1958) and Fienberg (1972) can be used in conjunction with the above results to produce estimators for the asymptotic variances of the $\hat{n}_{i;22\ldots2}$, $\hat{N}_i$ and $\hat{N}$. In particular, these quantities reduce to

$$V_{\hat{n}_{i,22\ldots2}} = n_i^2 V_{\hat{\gamma}_i} + \{\hat{n}_{i;22\ldots2}^3/n_i \hat{N}_i\}, \quad (3.7)$$

$$V_{\hat{N}_i} = n_i^2 V_{\hat{\gamma}_i} + \{\hat{n}_{i;22\ldots2}\hat{N}_i/n_i\}. \quad (3.8)$$

$$V_{\hat{N}} = \sum_{i=1}^{s}\sum_{i'=1}^{s} n_i n_{i'} V_{\hat{\gamma}_i,\hat{\gamma}_{i'}} + \sum_{i=1}^{s}\{\hat{n}_{i;22\ldots2}\hat{N}_i/n_i\} \quad (3.9)$$

where $V_{\hat{\gamma}_i}$ and $V_{\hat{\gamma}_i,\hat{\gamma}_{i'}}$ are estimates for the variance of $\hat{\gamma}_i$ and covariance of $\hat{\gamma}_i$ and $\hat{\gamma}_{i'}$ from (3.6).

For the case $s = 1$, the estimators for the asymptotic variances of $\hat{n}_{i;22\ldots2}$ and $\hat{N}_i$ are essentially the same as those given by Fienberg (1972) but avoid iterative computations required in general by his approach even after estimation of $\hat{\gamma}_i$, $\hat{n}_{i;22\ldots2}$ and $\hat{N}_i$.

4. Strategies for Fitting Log-Linear Models

In this section, we describe three strategies available for estimating parameters and fitted joint detection probabilities for the models of Section 2. Associated statistics for evaluating adequacy of fit are also referenced.

The most general method involves a slight modification of the approach of Grizzle and Williams (1972) for fitting log-linear models to complete contingency tables, which they developed as an application of the general methodology described by Grizzle, Starmer and Koch (1969). Weighted least-squares (WLS) computational algorithms are applied to fit the postulated log-linear model to the observed vector $\log p$, where p contains the various $p_{i;j_1j_2\cdots j_d}$, the unrestricted maximum likelihood estimate of $\pi_{i;j_1j_2\cdots j_d}$. The covariance matrix used is obtained by substituting the $p_{i;j_1j_2\cdots j_d}$ for $\pi_{i;j_1j_2\cdots j_d}$ in the asymptotic covariance matrix of $\log p$, determined by applying the δ-method for deriving the covariances of transformed random variables (see Grizzle, Starmer and Koch (1969)). Thus, $\log p$ is expanded in a Taylor series about $\log \pi_0$, and the covariance matrix of the linear term extracted. This method yields a direct estimate $\bar{\beta}$ of β without iteration; the fitted joint detection probabilities are obtained by substituting $\bar{\beta}$ into the model equations. The estimator $\bar{\beta}$ is a member of the class of procedures based on minimizing Neyman's (1949) modified chi-square criterion subject to a linearized hypothesis. As such, it is a Best Asymptotically Normal (BAN) estimate of β. For moderate to large samples in practice, $\bar{\beta}$ tends to be close to the estimate $\tilde{\beta}$ which maximizes the conditional likelihood based on the n_i, in the sense that individual components of $\bar{\beta}$ and $\tilde{\beta}$ tend to differ by less than the estimated standard deviation of either. The usual weighted least-squares algorithms produce test statistics, both for fit of the model and additional linear parametric restrictions, which belong to the class of test criteria defined by Wald (1943). Thus, in terms of asymptotic distribution and power, they are equivalent to the corresponding likelihood ratio tests based on the conditional likelihood. All computations for this approach may be executed using a general computer program for the analysis of categorical data, GENCAT (Landis, Stanish, Freeman and Koch (1976)), available from the University of Michigan.

When the postulated model is hierarchical, in terms of the entire set of dimensions involving both the different recording systems and the stratification variables, then the maximum likelihood estimators and likelihood ratio tests (based on the conditional likelihood) may be easily obtained. In this situation

$$[X^*] = \bigcup_{m=1}^{M} [X^*_{(m)}]$$

where $X^*_{(m)}$ is the design matrix of a factorial model involving a subset of the dimensions determined by the recording systems and/or stratification variables. If M is minimal, the observed marginal tables generating the $X^*_{(m)}$, $m=1,2,\ldots,M$, form a set of minimal sufficient statistics for the parameters of the model X^* (Birch (1963), Bishop, Fienberg and Holland (1975)). The sufficient statistics not only generate the fitted table which maximizes the underlying likelihood, but in fact are reproduced by it, as the maximum likelihood estimates of joint detection probabilities are the unique set of probabilities which both satisfy the model structure and generate marginal expected counts identical to the set of minimal sufficient statistics (Birch (1963)). This result is expressed in more general terms by expression (3.3).

For some models, the fitted joint detection probabilities may be calculated explicitly and directly from the sufficient marginal tables. Generally, however, it is necessary to use a modification of the technique of "iterative proportional fitting" (IPF, or "raking") of Deming and Stephan (1940), which converges correctly in all cases. If $C_{(m)}$ is the observed marginal table generated by $X^*_{(m)}$, the technique is executed as follows:

i) form Table T, of the same dimensions as the observed incomplete table, with zeros in cells representing unobserved elements and units in all other cells.

ii) collapse T to form the marginal array $C^{(1)}_{(1)}$ generated by $X^*_{(1)}$ from T; form $T^{(1)}_{(1)}$ by inflating each cell of T by the ratio of its marginal category frequency in $C_{(1)}$ to that in $C^{(1)}_{(1)}$.

iii) for $m = 2,\ldots,M$ form $C^1_{(m)}$ from $T^1_{(m-1)}$ using $X^*_{(m)}$; form $T^1_{(m)}$ from $T^1_{(m-1)}$ using $C^1_{(m)}$.

iv) for $\nu \le 2$, cycle through ii) to iii) substituting $T^{(\nu-1)}_{(m)}$ for T, $C^{(\nu)}_{(m)}$ for $C^{(1)}_{(m)}$, and $T^{(\nu)}_{(m)}$ for $T^{(1)}_{(m)}$; continue until $T^{(\nu)}_{(m)}$, and $T^{(\nu-1)}_{(m)}$ are sufficiently close.

The elements of $T^{(\nu)}_{(m)}$ are then divided by the appropriate stratum sizes n_i to yield the joint detection probabilities for each stratum. The estimated parameter vector $\tilde{\beta}$ is obtained by substituting these into the model equations and solving, if desired. Estimated population size for each stratum, and likelihood ratio tests of fit associated with the model, or with comparisons of alternate models, may be calculated using the fitted probabilities without ever explicitly obtaining the model parameters.

The estimates obtained by IPF may be preferable to those given by the WLS procedure when some observed stratum counts are modest, inasmuch as the asymptotic theory for the maximum likelihood estimators depends on the expected counts in cells of the marginal tables $C_{(m)}$, $m = 1,2,\ldots,M$, where as that underlying the WLS approach depends on expected counts in individual cells of the incomplete table. All computations for the IPF-MLE analysis may be executed using a computer program for fitting log-linear models to contingency tables, ECTA, available from the University of Chicago.

When modest observed counts make the use of WLS unattractive and the proposed model is not hierarchical, estimates may often be derived by applying a Functional Asymptotic Regression Methodology (FARM) approach. This capitalizes on the observation that non hierarchical models can be written as hierarchical models with linear restrictions on the parameters. Thus, we attempt to find an unsaturated hierarchical model from which the non hierarchical model of interest may be derived through the imposition of linear restrictions. IPF is applied to derive an initial estimate of β whose sampling variability derives from the expected counts corresponding to margins of the observed data table rather than interior cells. WLS algorithms are then applied to this preliminary estimate of β, using its estimated asymptotic covariance matrix under the hierarchical model (as determined by the δ-method), to introduce the linear restrictions which reduce the hierarchical model to the more parsimonious non hierarchical model of interest. The appropriate likelihood ratio test is used to assess fit of the initial hierarachical model, and a conditional WLS test used to evaluate adequacy of the subsequent reduction. Further reductions of the non hierarchical model may be evaluated by the application of WLS to the fitted parameters.

The FARM procedure is somewhat simpler to implement computationally than to describe conceptually. Once an initial hierarchical model is chosen, fitted joint detection probabilities are obtained under this model by IPF or WLS. The FARM estimate of β is then obtained by applying the WLS computational algorithms to the vector of these estimated probabilities instead of the observed proporation vector p. As a result, the FARM analysis may be performed, when IPF is used in the first stage, by simple execution in sequence of the computer programs ECTA and GENCAT described previously.

The literature describing development of log-linear model theory and the fitting strategies described here is vast, and no attempt has been made in this paper to adequately credit the contributors. The application of log-linear model theory to the MRS-CMR problem is due to Fienberg (1972), and a full exposition of the IPF-MLE approach appears in Bishop, Fienberg and Holland (1975). The WLS approach was adapted and described by Koch, El-Khorazaty and Lewis (1977), while FARM procedures are due to Koch, Imrey, Freeman, and Tolley (1977), and are applied to the MRS problem by El-Khorazaty, Imrey, Koch and Lewis (1977).

ACKNOWLEDGEMENTS

This research was in part supported by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 76-93. The authors would like to thank Jackie O'Neal for her conscientious typing of this manuscript.

REFERENCES

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society*, Series B, 25, 220-33.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Mass.

Chakraborty, P.N. (1963). On a method of estimating birth and death rates from several agencies. *Calcutta Statistical Association, Bulletin 12*, 106-12.

Chandrasekar, C. and Deming, W.E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, 44, 101-15.

Darroch, J.N. (1958). The multiple-recapture census, I. Estimation of a closed population. *Biometrika*, 45, 343-59.

Das Gupta, P. (1964). On the estimation of the total number of events and the probabilities of detecting an event from information supplied by several agencies. *Calcutta Statistical Association, Bulletin 13*, 89-100.

Deming, W.E. and Stephan, F.F. (1940). On a least-square adjustment of a sampled frequency table when the marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-44.

El-Khorazaty, M.N. (1975). Methodological strategies for the analysis of categorical data from multiple-record systems. University of North Carolina Institute of Statistics Mimeo Series No. 1019, Chapel Hill, N.C.

El-Khorazaty, M.N., Imrey, P.B., Koch, G.G. and Lewis, A.L. (1977). Applications of functional asymptotic regression methodology to multiple-record system data. In preparation.

Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, 59, 591-603.

Forthofer, R.N. and Koch G.G. (1973). An analysis for compounded functions of categorical data. *Biometrics*, 29, 143-57.

Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrics*, 25, 489-504.

Grizzle, J.E. and Williams, O.D. (1972). Log-linear models and tests of independence for contingency tables. *Biometrics*, 28, 137-56.

Koch, G.G., El-Khorazaty, M.N. and Lewis, A.L. (1977). The asymptotic covariance structure of log-linear model estimated parameters for the multiple recapture census. To appear in *Communications in Statistics A*.

Koch, G.G., Freeman, D.H., Jr., Imrey, P.B. and Tolley, H.D. (1977). The asymptotic covariance structure of estimated parameters from contingency table log-linear models. To appear in *Biometrics*.

Landis, J.R., Stanish, W.M., Freeman D.H., Jr., and Koch, G.G. (1976). A computer program for the generalized chi-square analysis of categorical data using weighted least-squares to generate Wald statistics. To appear in *Computer Programs in Biomedicine*.

Neyman, J. (1949). Contributions to the theory of the chi-square test. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, 230-73. University of California Press, Berkeley, Calif.

Wald, A. (1943). Tests of statistical hypotheses concerning general parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426-82.

A MODEL FOR AN INTEGRATED INFORMATION AND REFERRAL NETWORK FOR HUMAN SERVICES DELIVERY SYSTEMS

Arthur J. Ellick and Scott Byrd
Community Service Council of Broward County, Inc.

Information and Referral has long been regarded only as a link between needy persons and service providers. In today's climate of increased accountability and with a greater significance being given to cost-benefit analysis as it can be applied to the area of human services, information and referral should be viewed as a base from which data can be gathered on the use and availability of specified services. With this goal in mind, the development of a model of an I&R system was undertaken; however, since there is usually more than one agency in any geographic area which provides I&R, it became obvious that what was needed was an integrated network of I&R service providers which would form the nucleus of a data gathering system. Thus, the goal of devising an integrated system composed of all I&R providers forming a cohesive mechanism from which reliable data could be accumulated.

Obviously, the task of data gathering would have been much simpler if there were only one unique agency in any geographic area providing I&R however, since this is not the case, it is necessary to reach a common understanding among local I&R providers. The key to reaching this goal is the maintenance and preservation of individual sovereignty. Agencies must not be given the indication that the consolidation or elimination of individual I&R providers will ever occur. With these thoughts in mind, we can begin discussion of the integrated I&R network.

Prior to the initial planning of the network, consensus must be reached in two distinct areas. First, the "Community Planners" must reach a decision as to the type of data they wish to acquire. Secondly, it is essential that the "Community Planners" designate one local I&R provider as the central coordinator of the Information and Referral Network. Upon reaching consensus on these two points, only one major task remains; that of establishing a reference source of local service providers. This must include essential data such as location of service providers, hours of service, eligibility requirements (if any), and other information necessary for the provision of quality I&R. Once this reference file is complete, the establishment of the integrated I&R network can begin.

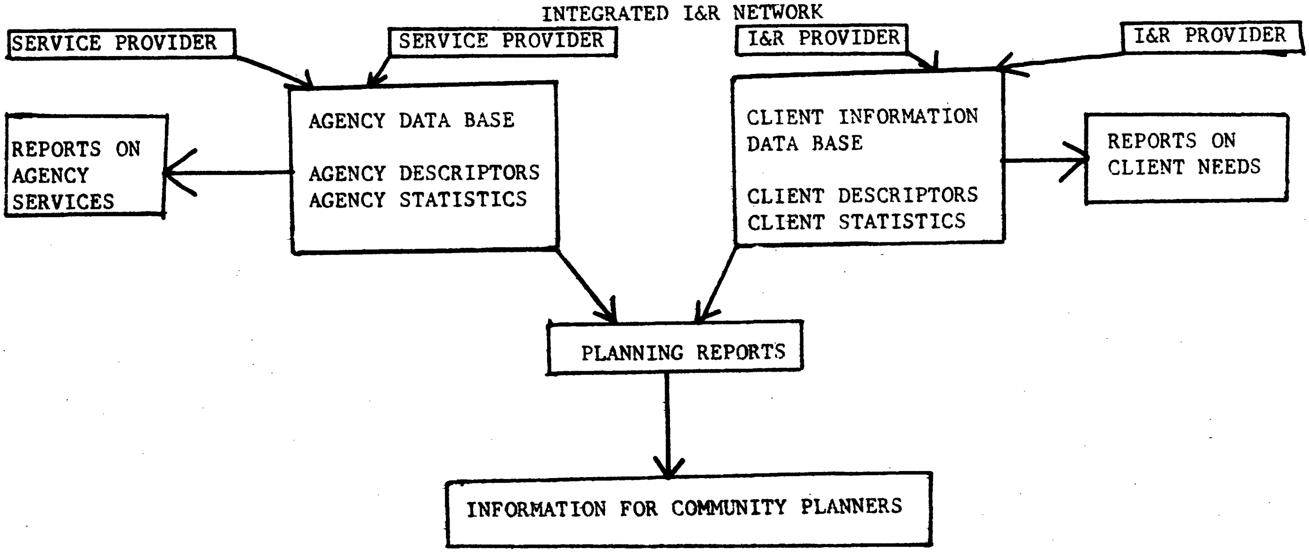
The model proposed by this paper consists of two distinct, but highly related, computerized sub-systems (Diagram A presents the model in descriptive terms). The first system is known as the agency data bank sub-system. This component of the system consists of descriptive data on each of the agencies stored within the computer memory. The data is stored in a structured manner so as to facilitate access of agency data. The model must provide for access of data by several varying classifications. Through the use of numerical codes and structural storage, data may be accessed through several varying means. Requests may be made for a listing of all agencies serving the aged, or who serve youth. In similar fashion, a listing may be obtained for: 1) all agencies in a geographic sector, or 2) those agencies serving Medicaid patients, or 3) those agencies with-in eligibility requirements, or 4) any other characteristic which is common to a group of service providers. In this manner, an information and referral agency equipped with a computer terminal connected to the proposed model can obtain specific data on where to refer a specific need for assistance. With this component of the model inaccurate referrals can be totally eliminated. For instance, a person requiring emergency food approaches an I&R agency and requests assistance. Instead of a counselor using information which exists in an out-dated fashion, the I&R counselor can access the computer data and obtain accurate information ensuring an efficient and effective referral.

The second sub-system is known as the client information data base. When an I&R counselor accepts an intake from a client, a client data form is completed. This form may exist in either of two fashions. It may exist in a printed form with answers to statistical questions recorded on paper and fed into the computer at a later date. The second and more efficient fashion is for data on the client to be directly fed into the computer at the time of original client contact. In this way, the computer, with the specific needs of the client already stored in memory can link up the stated needs with service providers who meet certain specific criteria. The services needed, the location and hours offered, the fees to be charged, and other critical factors can be matched with similar data in the agency data bank sub-system. The computer can automatically search out the "best match" and can thus make the most effective referral. As always, there are cases of need for which no services are offered. In this case, the computer stores this fact in memory. With the current capability of modern computer systems, it is no problem for the system to handle several inquires at once. Each I&R agency participating in the system is assigned a code unique to that agency. The computer is programmed to store data on each coded agency for retrieval at a later date.

Periodic reports on the activities of each I&R agency can be created from stored data. Like wise, the problem of clients seeking similar assistance from two or more sources is eliminated. Each agency connected to the computer is able to check past records for service given to a specific client through the use of a specified client number. This could be a social security number or some combination of name or initials unique for each client. In order to ensure confidentiality, real names are protected from public access and are available only to the agency initially filing the computerized report.

The major benefit of this proposed system is in the availability of data gathered over time on needs and services available. After the initial operation of several months to one year, sufficient data would be available to "Community Planners" for the determination of where the duplication of services un-met needs exist. With this data available, responsible leaders in the field of human service can push for the elimination of expensive non-efficient services which might duplicate those provided elsewhere at lower cost. In addition, the needs of the community which cannot be served by existing local services can now be clearly distinquished and services to meet these needs can be created. Finally, cost-benefit studies of service providers can be initiated with the long-range goal of the elimination of inefficient agencies. While this may sound harsh to many, the decline of human service dollars which is expected to occur in the future will mandate that only those services which are cost efficient and provide quality needed services will remain. The economics of sound business management can, and must be accepted into the realm of human services as they exist today.

Diagram A
INTEGRATED I&R NETWORK



WEIGHTED LEAST-SQUARES VERSUS MAXIMUM LIKELIHOOD ESTIMATES OF NON-HIERARCHICAL MODELS

Mark Evers, Duke University
N. Krishnan Namboodiri, University of North Carolina at Chapel Hill

In this paper we analyze two sets of categorical data with the following objectives: (1) to show that fitting non-hierarchical models is very feasible and that such models may give extremely good fit, when hierarchical models may not, and (2) to assess the relative utility of maximum likelihood (ML) and weighted least-squares (WLS) estimation techniques in light of data structure and available computer programs. We assume familiarity with the basic hierarchical log-linear technique as presented, for example, in Goodman [5, 6, 7].

Illustration 1: Self-esteem data

The data in Table 1, taken from Rosenberg [11], show proportions of persons with "high" self-esteem by religion (three categories) and father's education (six categories). We take self-esteem to be a dichotomous variable ("high" or not "high") and treat it as the dependent variable. When using the ML hierarchical method [6] in this data situation, one typically starts by testing whether father's education and religion interact in their effects on self-esteem, and, if that three-way interaction is absent, the next step would be to inquire whether the main effects of father's education and religion are significant. The first step, then, is to fit the three two-way marginals, which is equivalent to "fitting" the hypothesis of no three-way interaction [3]. In this case, such a model fits the data very poorly ($\chi^2 = 37.66$, 10 df, $p < .001$), and the investigator would infer that religion and father's education do indeed interact in their effects on self-esteem. At this point the typical log-linear analysis would stop, for the only model among the hierarchical ones that would give a better fit would be the so-called "saturated model," which, because it uses all the degrees of freedom available, yields no data reduction whatsoever, enabling the analyst to do no more than describe the observed frequencies in the table.

We shall now briefly describe how one may try to identify a parsimonious non-hierarchical model that fits the data extremely well. We first fit the saturated model, using a program called NONMET which is a WLS routine [8]. (For documentation, write to the Institute for Research in Social Science, University of North Carolina, Chapel Hill.) In order to fit any model using the NONMET program, we must first specify the design matrix using the so-called "effect coding" [10, pp. 121-128]. Given below are the first eight columns of the design matrix. These columns represent the "general mean" and the "main effects" of religion (two components, R_1 and R_2) and father's education (five components, E_1, ..., E_5). The ten remaining columns in the design matrix (not shown) correspond to the interaction effects. They can be derived by multiplying corresponding elements of one column (R_1 or R_2) for religion and one (E_1, E_2, E_3,

Row	Columns of Design Matrix							
	(1) $\bar{X}$	(2) R_1	(3) R_2	(4) E_1	(5) E_2	(6) E_3	(7) E_4	(8) E_5
C:1	1	1	0	1	0	0	0	0
2	1	1	0	0	1	0	0	0
3	1	1	0	0	0	1	0	0
4	1	1	0	0	0	0	1	0
5	1	1	0	0	0	0	0	1
6	1	1	0	-1	-1	-1	-1	-1
J:1	1	0	1	1	0	0	0	0
2	1	0	1	0	1	0	0	0
3	1	0	1	0	0	1	0	0
4	1	0	1	0	0	0	1	0
5	1	0	1	0	0	0	0	1
6	1	0	1	-1	-1	-1	-1	-1
P:1	1	-1	-1	1	0	0	0	0
2	1	-1	-1	0	1	0	0	0
3	1	-1	-1	0	0	1	0	0
4	1	-1	-1	0	0	0	1	0
5	1	-1	-1	0	0	0	0	1
6	1	-1	-1	-1	-1	-1	-1	-1

E_4, or E_5) for father's education. This 18 by 18 design matrix when used in NONMET yields estimates of the parameters of the saturated model.

Note that the WLS routine programmed in NONMET predicts the proportion of persons with "high" self-esteem or the logarithm or the logit thereof. Here we confine ourselves to the logit form. (Logit is defined as the natural logarithm of $p/(1-p)$, where p is the proportion of cases with "high" self-esteem.)

With the ML model we are not predicting the logit of the dependent proportion, but rather the logarithm of the frequency, or the logarithm of the proportion of the total number of cases in each cell of the three-way table of religion-by-father's education-by-self-esteem [1, 5, 6]. (Let us call the ML model the log-frequency model.) Since, with this model we predict the cell frequencies rather than the logits, we are now predicting 36 values. Therefore, the design matrix needed to predict the observed frequencies in the three-way table will have 36 rows, one corresponding to each cell in the table. The matrix for the saturated log-frequency model will also have a total of 36 columns, for we must now estimate parameters representing the general mean (1), main effects (self-esteem, 1; religion, 2; father's education, 5), the two-way effects (self-esteem and religion, 2; self-esteem and father's education, 5; religion and father's education, 10) and the three-way effects (10). It is easy to demonstrate that the parameters involving self-esteem for the saturated log-frequency model are exactly half of the corresponding parameters of the logit model [6].

For the saturated model, the ML estimates can be easily obtained using a program such as ECTA. We can also use a more general ML estimation program, such as MAXLIK [9], and such a program must be used to obtain estimates for any model other than the hierarchical variety described by Goodman.

The WLS estimates and their standard errors for the saturated logit model described above are shown below:

R_1:	-.1108(.0636)	R_1E_3:	.3418(.1016)
R_2:	.3334(.0817)	R_1E_4:	-.0132(.1474)
E_1:	-.1900(.1200)	R_1E_5:	-.3531(.1556)
E_2:	-.0907(.0841)	R_2E_1:	-.1655(.2164)
E_3:	-.2569(.0815)	R_2E_2:	-.3214(.1609)
E_4:	.0592(.1115)	R_2E_3:	-.0366(.1395)
E_5:	.2403(.1169)	R_2E_4:	-.0348(.1799)
R_1E_1:	.1007(.1375)	R_2E_5:	-.4508(.1975)
R_1E_2:	.0202(.1042)		

(It can be easily shown that for the saturated model the WLS estimates and the ML estimates are identical.)

Each estimate shown above can be used to test the significance of the corresponding parameter by calculating the statistic (estimate/standard error)^2, which is distributed asymptotically as chi-square with one degree of freedom. This procedure yields the following parameters as "significant": R_1, R_2, E_1, E_3, E_5, R_1E_3, R_1E_5, R_2E_2, and R_2E_5. Fitting a model containing only these parameters and the grand mean may provide a parsimonious representation of the data in question. Such a model would leave 8 degrees of freedom, since it fits only 10 of the 18 parameters of the saturated model. However, in this case we can further reduce the number of parameters that need to be estimated by fitting (E_1 - E_5) instead of the pair E_1 and E_5, and $(R_1-R_2)E_5$ instead of the pair R_1E_5 and R_2E_5. With this further reduction, we estimate a total of 8 parameters, leaving 10 degrees of freedom. The design matrix for this last model is shown below:

Row	(1) X	(2) R_1	(3) R_2	(4) E_1-E_5	(5) E_3	(6) $(R_1-R_2)E_5$	(7) R_1E_3	(8) R_2E_2
	Columns of Design Matrix							
C:1	1	1	0	1	0	0	0	0
2	1	1	0	0	0	0	0	0
3	1	1	0	0	1	0	1	0
4	1	1	0	0	0	0	0	0
5	1	1	0	-1	0	1	0	0
6	1	1	0	0	-1	-1	-1	0
J:1	1	0	1	1	0	0	0	0
2	1	0	1	0	0	0	0	1
3	1	0	1	0	1	0	0	0
4	1	0	1	0	0	0	0	0
5	1	0	1	-1	0	-1	0	0
6	1	0	1	0	-1	1	0	-1
P:1	1	-1	-1	1	0	0	0	0
2	1	-1	-1	0	0	0	0	-1
3	1	-1	-1	0	1	0	-1	0
4	1	-1	-1	0	0	0	0	0
5	1	-1	-1	-1	0	0	0	0
6	1	-1	-1	0	-1	0	1	1

The WLS estimates of the parameters of this model and their standard errors are:

R_1:	-.1112(.0564)	$(R_1-R_2)E_5$:	-.3052(.1325)
R_2:	.3344(.0743)	R_1E_3:	.3606(.0709)
E_1- E_5:	-.1435(.0692)	R_2E_2:	-.2839(.0933)
E_3:	.2565(.0649)		

The corresponding ML estimates obtained by using the MAXLIK program are:

R_1:	-.1124	$(R_1-R_2)E_5$:	-.3096
R_2:	.3370	R_1E_3:	.3604
E_1- E_5:	-.1444		
E_3:	-.2572	R_2E_2:	-.2846

The goodness of fit chi-square for the WLS procedure is 3.15 (10df, p. = .978) and that for the ML procedure is 3.18 (10df, p = .977).

Given the near identity of the estimates obtained using the two techniques, the analyst would be advised to employ the technique that is easiest to use, and less expensive. Since the ML programs such as MAXLIK are quite costly and require the input of initial estimates, which must be obtained from prior calculations, it seems clear that for data where we can specify one of the variables as dependent and where that variable is dichotomous, it is preferable to use the WLS technique. It deserves to be emphasized that the analysis procedure just described makes sense only if all the parameters can be given substantive interpretations. (The theoretical significance of these particular parameters, of course, will have to be found on the basis of non-statistical considerations that are beyond the scope of this paper.) We wish merely to illustrate that these eight parameters suffice to describe virtually all of the variation in self-esteem in the given data set.

Illustration 2: Openness to Change

The second illustration uses data in which the dependent variable is a polychotomy. The data are from a survey reported in Duncan [2].

In that survey, respondents were asked about their attitudes toward "making changes in the way our country is run" [2, pp. 177-181]. Four response categories were used: R_1: "We should rarely, if ever, make changes", R_2: "We should be very cautious in making changes", R_3: "We should feel free to make changes", R_4: "We must constantly make changes." Suppose we are interested in assessing the effect of the year (Y) of the survey (1956 or 1971) and the respondent's political party identification (P_1: Republican, P_2: Democrat, or P_3: Independent).

Analyzing these data using the WLS technique to "search" for a parsimonious model is not as straightforward as in the previous illustration, because the dependent variable is now polychotomous. We start by viewing the response-by-party-by-year distribution as a multinomial (with 24 classes) and let the multinomial proportion in each cell take the place of the binomial proportion in Illustration 1, treating log p as the quantity that is predicted. It is important to note that the procedure is statistically incorrect and may produce bias in our estimates of the variance-covariance matrix. Nevertheless, as we demonstrate below, this procedure is less problematic than it might appear at first blush. (We employ the WLS procedure only for a preliminary analysis, which we follow up with the ML procedure. In the present case the ML estimates turn out to be substantially similar to the WLS estimates.)

Our strategy involves first fitting the saturated model to the data in Table 2. By omitting non-significant effects, specifying "difference" effects for pairs of related effects, and taking other similar steps, we obtain, after a few preliminary runs using NONMET, a parsimonious model that seems to fit the data extremely well. This model is then fitted using the MAXLIK program. The results are reported in Table 3.

In the analysis of these same data, Duncan [2] uses standard hierarchical ML procedures to arrive at a parsimonious and good-fitting model, by treating each polychotomous variable with g categories as a set of g dichotomous variables, in each case contrasting a given category with the $g-1$ remaining categories. He terms these dichotomous variables "formal" variables.

Duncan begins the analysis of these data by fitting a baseline model which posits independence between response and the joint variable year-by-party. He then proceeds to fit several models one at a time, each of which includes the parameters of the baseline model as well as one additional parameter representing the effect on one of the response categories of one of the independent variable categories. Thus, he shows twelve effects for party and response and four effects for year and response. The statistical significance of each additional parameter is ascertained by comparing the difference in chi-square of each model and the baseline model. Each of the twelve three-way interactions are also tested for significance one at a time, by comparing a model that included it with a model that excluded it. In this way, Duncan arrives at what we will for convenience term his "best-fitting" model, i.e., his Model (15), with 15 parameters, nine degrees of freedom, and a chi-square value of 3.5, with a probability exceeding .9.

Duncan's "best-fitting" model can be specified in a design matrix format: First think of each row of the matrix as being defined by the eight dichotomous formal variables, rather than by the underlying three variables. Then, for each of the formal variable effects, one can specify a column with a 1 for each cell corresponding to the category of the formal variable, and a -1 for all other cells. These are the so-called elementary column specifications shown in columns (1) through (8) in Table 4. These elementary columns can in turn be used to specify any of the joint (interaction) effects in the model. Thus, for example, column (9) shows the vector corresponding to the parameter R_4Y, which is obtained by multiplying corresponding elements in elementary columns (4) and (8). Columns (10) and (11) of Table 4 show two other interaction vectors that are specified according to the formal variable approach. The formal approach is used only to define the interaction terms in the model--terms involving the multiplication of two or more elementary vectors. It is important to note that all formal variable interaction terms involving a given polychotomous variable cannot be simultaneously included in the design matrix, because any one column for the g-category polychotomous variable would then be linearly dependent on the other $g-1$ columns. Likewise, we cannot use the vectors in the first eight columns of Table 4 to specify the main effects. Because of this, we resort to the more usual column specification for fitting main effects and specify all interaction effects in terms of formal variables in the manner described above and shown in Table 4.

Model (1) in Table 5 corresponds to Duncan's best-fitting model. We arrive at the design matrix specification for this model by specifying baseline model parameters according to the usual column design, with the additional interaction effects specified according to the formal variable design. Note that this design matrix specification yields results identical to Duncan's best-fitting model, and it allows for estimation of parameters without resorting to procedures for handling structural zeros.

The second model in Table 5 is a non-hierarchical model fitted to the same data (Table 2). Once a matrix approach is adopted, there is no need for the investigator to adhere only to hierarchical models. Note that the terms in model (2) are quite similar to those in model (1). Model (2) differs from model (1) in that the former excludes the effects P_1 and P_2Y and includes the effects R_1P_2Y and R_3P_2Y. This substitution of two parameters leaves unchanged the degrees of freedom but improves the fit of the model. In addition, the substantive inferences made on the basis of model (2) would differ from those made with model (1), corresponding to the two different parameters in the models.

It is interesting to note that not only do models (1) and (2) in Table 5 allow us to make slightly different inferences about the relationships in the data in Table 2, but also that both of these models differ substantially in structure from the parsimonious and slightly better fitting model in Table 3. Neither the different parameterization of the design matrix in our approach nor the introduction of the formal variable format accounts for any of the differences in structure between the model in Table 3 and the models in Table 5, since if the design matrix for the model in Table 3 were recast in terms of the formal variables there would be no change in the parameter estimates or goodness-of-fit statistic. We may remind the reader of Goodman's caution [4, p. 48] that different selection procedures may lead to different models, all of which concisely fit a given set of data. Another analyst may obtain yet another model that fits these data well.

Nevertheless, we wish to underscore the central point that any analyst is well-advised to consider non-hierarchical models as well as those that are hierarchical. Duncan's analysis of these data was restricted to hierarchical models, and was somewhat cumbersome in requiring that many models be fitted in order to arrive at one that was reasonably parsimonious.

Our initial analysis of the data in Table 2 yielded in straightforward fashion a very parsimonious model that was non-hierarchical. Reanalysis of the data using the formal variable approach with a design matrix allowed us to duplicate Duncan's finding, and then go on to fit a non-hierarchical model that describes the data somewhat better than does Duncan's model.

General Recommendations

(1) If the dependent variable is dichotomous, use a WLS program such as NONMET, rather than a hierarchical ML program such as ECTA. This allows the analyst to use his or her imagination in finding models that are theoretically appropriate, parsimonious, and that fit the data very well. Although such models may occasionally be hierarchical, in which case ECTA may be useful, a WLS program such as NONMET can be used to fit all the models fitted by ECTA and more. A program such as MAXLIK will then allow the analyst to obtain ML estimates, once the appropriate model has been found with the WLS technique.

(2) If the dependent variable is polychotomous, first apply a WLS program (e.g., NONMET) to the logarithm of cell frequencies to find the most suitable model as in the second illustration. Since the NONMET program in this case is statistically less attractive because it employs the wrong variance-covariance matrix, use a program such as MAXLIK to obtain ML estimates for the model.

(3) For a logit model, in which the dependent variable is dichotomous, it would be enough to use the NONMET program alone, without finding ML estimates, because both procedures yield virtually identical results.

(4) For a set of data with a polychotomous dependent variable, it may be the case that the use of the NONMET program in the manner we suggest above, although statistically not quite attractive, produces acceptable results for most analyses. For our example, the WLS estimate did not depart significantly from the ML estimates.

REFERENCES

[1] Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. Cambridge, Mass.: MIT Press.

[2] Duncan, Otis Dudley. 1975. "Partitioning Polytomous Variables in Multiway Contingency Analysis." Social Science Research 4 (September): 167-182.

[3] Goodman, Leo A. 1970. "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications." Journal of the American Statistical Association 65 (March): 226-256.

[4] ________. 1971. "The Analyses of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications." Technometrics 13 (February): 33-61.

[5] ________. 1972a. "A General Model for the Analysis of Surveys." American Journal of Sociology 77 (May): 1035-1086.

[6] ________. 1972b. "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables." American Sociological Review 37 (February): 28-46.

[7] ________. 1973. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." American Journal of Sociology 78 (March): 1135-1191.

[8] Grizzle, James E., C. Frank Starmer, and Gary C. Koch. 1969. "Analysis of Categorical Data by Linear Models." Biometrics 25 (September): 489-504.

[9] Kaplan, Ellen B. and R. C. Elston. 1972. "A Subroutine Package for Maximum Likelihood Estimation (MAXLIK)." Institute of Statistics Mimeo Series, No. 823. Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill.

[10] Kerlinger, Fred N. and Elazur J. Pedhazur. 1973. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston.

[11] Rosenberg, Morris. 1962. "Test Factor Standardization as a Method of Interpretation." Social Forces 41 (October): 53-61.

ACKNOWLEDGEMENT

We are grateful to Sharon Poss for extensive programming support, in particular for modifying MAXLIK and NONMET for the problems we address, to Eugenia Hatley for clerical assistance, and to Valerie Hawkins for typing the manuscript. The research was supported in part by grant SOC76-02100 from the National Science Foundation.

TABLE 1

PROPORTION WITH HIGH SELF ESTEEM, BY RELIGIOUS BACKGROUND AND FATHER'S EDUCATION

Religion*	Father's Education					
	Eighth grade or less	Some high schl.	H.S. grad.	Some col.	Col. grad.	Post-grad.
C:	.681	.685	.717	.709	.675	.729
(n)	(360)	(482)	(541)	(141)	(114)	(70)
J:	.718	.706	.745	.788	.879	.827
(n)	(39)	(126)	(137)	(85)	(99)	(75)
P:	.648	.720	.525	.699	.706	.738
(n)	(193)	(325)	(406)	(156)	(279)	(122)

*C: Catholic; J: Jewish; P: Protestant

TABLE 2

PERCENTAGE DISTRIBUTIONS OF RESPONSES TO "CHANGE" QUESTION, BY YEAR AND PARTY

Year & Response	Party: Rep. (P_1)	Dem. (P_2)	Ind. (P_3)
1956 (Y_1)			
R_1	2.0	2.1	1.1
R_2	47.0	49.0	51.1
R_3	37.0	38.1	31.1
R_4	14.0	10.9	16.7
Total	100.0	100.0	100.0
(n)	(200)	(431)	(90)
1971 (Y_2)			
R_1	1.3	2.2	0.4
R_2	53.4	42.4	36.0
R_3	25.8	32.6	33.0
R_4	19.5	22.8	30.6
Total	100.0	100.0	100.0
(n)	(159)	(509)	(242)

TABLE 3

ESTIMATES FOR PARSIMONIOUS MODEL FITTED TO DATA IN TABLE 6

Parameter	WLS Estimate	ML Estimate
R_1	-2.2824	-2.2958
R_2	1.1862	1.1910
R_3	.8513	.8540
P_1	- .2710	- .2728
P_2	.7779	.7816
Y	.0960	.0961
P_1Y	- .3009	- .3022
P_2Y	.0521	.0510
R_1P_2	.2944	.3022
R_1Y	- .3220	- .3196
R_1P_2Y	.2901	.2837
$R_2(P_1-P_2)$	.0847	.0875
$R_2(P_1-P_2)Y$	.1326	.1315
R_3P_2Y	- .1529	- .1519
Chi-square		
value	2.5720	1.6214
df	9	9
p	.9789	.9774

TABLE 4

COLUMN SPECIFICATION FOR DESIGN MATRIX CORRESPONDING TO FORMAL VARIABLE APPROACH

								Elementary Column Specifications								Interaction Columns		
								(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
R_1	R_2	R_3	R_4	P_1	P_2	P_3	Y	R_1	R_2	R_3	R_4	P_1	P_2	P_3	Y	R_4Y	R_1P_3	R_2P_1Y
1	2	2	2	1	2	2	1	1	-1	-1	-1	1	-1	-1	1	-1	-1	-1
1	2	2	2	1	2	2	2	1	-1	-1	-1	1	-1	-1	-1	1	-1	1
1	2	2	2	2	1	2	1	1	-1	-1	-1	-1	1	-1	1	-1	-1	1
1	2	2	2	2	1	2	2	1	-1	-1	-1	-1	1	-1	-1	1	-1	-1
1	2	2	2	2	2	1	1	1	-1	-1	-1	-1	-1	1	1	-1	1	1
1	2	2	2	2	2	1	2	1	-1	-1	-1	-1	-1	1	-1	1	1	-1
2	1	2	2	1	2	2	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
2	1	2	2	1	2	2	2	-1	1	-1	-1	1	-1	-1	-1	1	1	-1
2	1	2	2	2	1	2	1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1
2	1	2	2	2	1	2	2	-1	1	-1	-1	-1	1	-1	-1	1	1	1
2	1	2	2	2	2	1	1	-1	1	-1	-1	-1	-1	1	1	-1	-1	-1
2	1	2	2	2	2	1	2	-1	1	-1	-1	-1	-1	1	-1	1	-1	1
2	2	1	2	1	2	2	1	-1	-1	1	-1	1	-1	-1	1	-1	1	-1
2	2	1	2	1	2	2	2	-1	-1	1	-1	1	-1	-1	-1	1	1	1
2	2	1	2	2	1	2	1	-1	-1	1	-1	-1	1	-1	1	-1	1	1
2	2	1	2	2	1	2	2	-1	-1	1	-1	-1	1	-1	-1	1	1	-1
2	2	1	2	2	2	1	1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1
2	2	1	2	2	2	1	2	-1	-1	1	-1	-1	-1	1	-1	1	-1	-1
2	2	2	1	1	2	2	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1
2	2	2	1	1	2	2	2	-1	-1	-1	1	1	-1	-1	-1	-1	1	1
2	2	2	1	2	1	2	1	-1	-1	-1	1	-1	1	-1	1	1	1	1
2	2	2	1	2	1	2	2	-1	-1	-1	1	-1	1	-1	-1	-1	1	-1
2	2	2	1	2	2	1	1	-1	-1	-1	1	-1	-1	1	1	1	-1	1
2	2	2	1	2	2	1	2	-1	-1	-1	1	-1	-1	1	-1	-1	-1	-1

TABLE 5

ML ESTIMATES OF PARAMETERS AND GOODNESS OF FIT CHI SQUARES OF TWO PARSIMONIOUS MODELS FOR DATA IN TABLE 2: MODELS SPECIFIED USING FORMAL VARIABLE APPROACH

Parameter	Model 1	Model 2
R_1	-2.4839	-2.7211
R_2	1.2535	1.3323
R_3	.8991	.9763
P_1	- .1159	--- a
P_2	.8443	.9437
Y	.2729	.2534
P_1Y	- .2560	- .2873
P_2Y	- .0654	--- a
R_2Y	.0501	.0529
R_4Y	.1909	.1862
R_1P_3	- .2706	- .4309
R_2P_1	.0471	.0443
R_4P_3	.0929	.0960
R_1P_2Y	--- a	.0595
R_2P_1Y	.0737	.0891
R_3P_2Y	--- a	.0350
Chi-square		
value	3.477	2.643
df	9	9
p	.942	.977

a Parameter excluded from the model

EARNINGS DATA FROM THE CPS: NEW COLLECTION EFFORTS AND SOME FINDINGS

Paul O. Flaim, Bureau of Labor Statistics

The Bureau of Labor Statistics has traditionally relied--and, indeed, continues to rely--on establishment surveys as the principal vehicles for the collection of earnings data.

However, establishment-based data generally tell us nothing about such important socioeconomic issues as the differences in earnings between blacks and whites, men and women, the young and the old, the highly educated and the less educated. Such information can simply not be collected from employers without imposing an unreasonable burden of record keeping and reporting upon them. The only practical way to collect such information is through household-type sample surveys, that is, through interviews with either the workers themselves or with some other member of their families.

The oldest and largest undertaking of this nature--excluding the Decennial Census--is the Current Population Survey (CPS), instituted in 1940 to measure the extent of labor force participation and unemployment among the American population. Over the years, this survey has been expanded several times and has been used more and more as a vehicle for the collection of other wide ranging information, including earnings data which could be linked to the characteristics of individual workers. 1/

Expanded collection efforts

Data on annual earnings (and income from other sources) have long been collected through the CPS. This information, which is obtained each March, refers to aggregate earnings (and other income) for the previous calendar year. These data have been a most important source of information on the historical trends in earnings of specific population groups. However, except where they apply to persons working full time the year round, they cannot be used as a precise tool for the measurement of intergroup differences in earnings, inasmuch as the extent of time worked, in terms of hours per week and weeks per year, is generally reported and recorded only in very broad intervals. In order to overcome these limitations, the Bureau of Labor Statistics has been aiming for the collection of CPS-based earnings data which would be both more current and linked to more precisely defined time periods.

The first step in this direction was taken in 1967, when a single question on "usual weekly earnings" was added to the questionnaire used annually in conjunction with the May survey. In 1973, the May questionnaire was expanded further, the principal addition being a question on "usual weekly hours" and one on the hourly wage rate for workers paid by the hour.

A glance at the findings

The data on "usual weekly earnings" have now been published for several years and they have shed considerable light on the earnings trends of specific group of workers. 2/ These annual data have shown that earnings for most groups of workers have risen somewhat faster over the 1967-1976 period than indicated by the overall average based on establishment data. Over this period, the establishment based average showed hardly any increase in constant dollar terms, as has also been the case for the overall average for all wage and salary workers based on the CPS data. Nevertheless, the disaggregated data from the CPS show that most groups of workers, particularly adult workers with full-time jobs, enjoyed an increase in real earnings averaging about 10 percent over this period. (See table 1.) This would indicate that the gradual change in the mix of workers, that is, the increase in the proportion accounted for by women and youths, who are generally concentrated in low paid jobs, has been a primary factor in holding down the overall earning average for all workers.

In terms of racial comparisons, the data on weekly earnings show that the earnings gap between blacks and whites, though still large, has narrowed considerably since the late 1960's. Specifically, real weekly earnings for blacks working full time showed a 21 percent increase over the 1967-76 period, whereas the average for whites shows only a 4 percent gain.

Going beyond these "macro" findings, BLS economists have turned to the micro data on weekly and hourly earnings--that is the tape entries for individual workers--to examine other questions, such as the earnings differential between unionized and nonunionized workers 3/ and the aggregate amount of earnings being lost because of unemployment. 4/

Hourly earnings

The CPS data on hourly earnings, available since 1973, also show considerable potential for the examination of several issues. Aside from their obvious use as a more exact measure of intergroup differences in earnings, they can, for example, provide valuable insights on the impact of changes in minimum wage legislation.

The impact of the 1974 change, for example, is clearly visible when one compares the May 1973 distribution of hourly earnings for workers paid by the hour with the distribution for May 1975 (see table 2). Although the earnings intervals used to tabulate these data are relatively wide, the distributional impact of the increase in minimum wages from $1.60 to $2.10 per hour stands out rather well. Moreover, the data for May 1976 appear to give a good indication of the further changes which went into effect last January, when the minimum was raised to $2.30 for most workers and to $2.20 for a portion not previously covered. Another interesting trend, which reflects the general upcreep in wages, is the rapid rise from 1973 to 1976 in the number of workers reported as earning from $6 to $8 an hour.

A more comprehensive measure of hourly earnings covering all wage and salary workers has also been derived by dividing their "usual weekly earnings" by their "usual weekly hours." The distribution of these earnings data, also based on the May survey is shown in table 3. As might be expected, since these data cover all wage and salary workers including professionals, supervisors, and managers, they yield medians and means which are substantially higher than those for hourly workers shown in table 2.

Accuracy of data

The collection of earnings data through a household survey implies the trading off of a certain amoung of accuracy in order to acquire other specific information on the characteristics of the wage earners. The obvious question is how much accuracy is being sacrificed.

The accuracy lost because of sampling variability can, at least in theory, be readily computed. The problem is that even if we knew what the amount of sampling variability is, we still would not know whether there might be any systematic bias attached to these numbers. What we do know is that, for various reasons, the nonresponse rate for the earnings question in the May survey is relatively high, having ranged from about 16 to 20 percent over the years.

The best way to determine how accurately the data on weekly and hourly earnings are being reported would be through a so-called "records check," that is by comparing these data, at least for a small sample of workers, with the actual payroll records maintained by their employers. The BLS and the Bureau of the Census have long recognized the desirability of conducting such a test and, have tentatively scheduled it for January 1977.

In the meantime, one way to gauge the accuracy of the earnings data derived from the CPS is by comparing them with similar data derived from the BLS survey of establishments. In terms of the largest industry groups where the two surveys cover universes which are at least roughly similar, though never quite the same, the data for May 1975 showed the following patterns of weekly earnings.

Except for the construction industry, where the two surveys differ the most in terms of the universe covered (the establishment survey being limited to "contract construction" whereas the CPS covers all construction activity), the averages for the other industry groups shown below are reasonably close.

Another way to compare the data from the two surveys is in terms of hourly earnings. Table 4 shows such comparisons for all "two-digit" manufacturing industries, with the data from the establishment survey having been adjusted to exclude the impact of any overtime premiums. Of course, there remain some differences in coverage. For example, these CPS data apply only to workers who report that they are paid by the hour, while the establishment data apply to all production and nonsupervisory workers. Nevertheless, despite this and other differences, the data from the two surveys are again reasonably close. As shown, mean earnings from the CPS survey fall generally short of those from the establishment survey, but the differences aren't that great and the pattern is not at all erratic. In sum, when the various measurement differences are taken into account, the CPS based data on hourly and weekly earnings compare rather favorably with the data derived from the much larger survey of establishments.

Further tests

Being reasonably satisfied with the reliability of the CPS data on weekly and hourly earnings, the BLS has been exploring the possibility of having it collected more than once a year, that is either monthly or quarterly. Prior to undertaking such a regular collection program--which can not, of course, be started without budget authorization and the necessary clearances--the Bureau wanted to make sure that the pattern of survey questions to be used for this purpose was the best that could be developed.

To this end a special experiment was conducted in conjunction with the CPS survey in November 1975. The principal purposes of this experiment were:

1. To test a procedure which would allow workers not paid at an hourly or weekly rate to report their earnings in the most applicable terms--i.e. daily rate, monthly rate, annual rate, piecework basis, etc. 5/

	Actual weekly earnings of production and nonsupervisory workers from Establishment Survey	Usual weekly earnings of all wage and salary workers from Current Population Survey	
	(Mean)	(Median)	(Mean)
Mining	$248	$246	$260
Construction	263	215	229
Manufacturing	185	186	201
Transportation and public utilities	226	223	227
Wholesale and retail trade	125	113	137
Finance insurance, and real estate	149	155	193

2. To determine what gains in reliability might be achieved by obtaining the earnings information directly from the workers rather than, second hand, from other members of their households.

3. To see whether it would be feasible to collect information on the additional earnings obtained by workers in terms of tips, commissions, bonuses, etc.

In terms of how workers are paid, as shown below, a little over one half were reported as paid by the hour. The proportions reporting that they were paid either at weekly, biweekly, monthly, or annual rates were roughly equal and together, accounted for four-tenths of the distribution, with the balance (about 7 percent) being divided up among several other categories (daily rate, piecework rate, commission basis, etc.)

Percent distribution of workers by how paid, November 1975

	Total	Male	Female
Total reporting	100.0	100.0	100.0
Hourly rate	54.2	52.3	56.8
Daily rate	1.3	.8	2.0
Weekly rate	10.1	10.7	9.2
Biweekly rate	8.3	8.5	8.0
Monthly rate	11.4	10.9	11.9
Annual rate	9.2	10.2	7.9
Piecework rate	1.2	1.0	1.4
Commission basis	1.9	2.3	1.4
Salary plus commission	.8	1.2	.1
Salary plus tips	.1	-	.1
Other way	1.6	2.0	1.1

A comparison of the data for workers who reported their own earnings (the so-called "designated respondents") with the data for workers whose earnings were reported by other members of the household ("proxy respondents") does not show any clear differences. In order to make these comparisons possible, the November 1975 test panel, consisting of two-thirds of the CPS sample, was divided into two equally representative parts. In one of these two parts--the C sample--the interviewers were instructed to make every reasonable effort to interview the actual wage earners, whereas in the other part--one-half of the so-called A sample--they could obtain the data from either the wage earner or any other responsible member of the household. The final outcome was one sample panel where the "designated respondents" accounted for about three-fourths of the interviews and another panel where the "designated respondents" accounted for only 45 percent of the interviews.

The reported median earnings based on the unweighted data for the two sample groups were as follows:

Median earnings by how paid

How paid	"C" sample (73 percent designated respondents)	"A" sample (45 percent designated respondents)	"C" sample "A" sample
Hourly rate	$ 3.41	$ 3.41	1.00
Daily rate	24.00	25.00	.96
Weekly rate	178.00	182.00	.98
Biweekly rate	426.00	445.00	.96
Monthly rate	829.00	841.00	.99
Yearly rate	13,182.00	13,567.00	.97
Piece work by week	118.00	126.00	.94
Commissions, by week	226.00	199.00	1.14
Salary plus commission, by week	260.00	285.00	.91
Salary plus tips, by week	62.00	62.00	1.00
Other way, by week	209.00	206.00	1.01

As shown above, for the majority of the workers, those paid by the hour, there was no difference between the median obtained from the "C" sample, where the designated respondents accounted for three-fourths of the responses, and that from the "A" sample, where most of the data was obtained from proxy respondents. For workers paid at other than hourly rate, the "C" sample yielded medians which, for the most part, were slightly lower than those from the "A" sample. The differences, however, were relatively small.

In sum, the two sample panels yielded very similar results. For those groups where the results were slightly different, it cannot be determined, short of a record check, which of two panels yielded the most accurate information. Although we are still inclined to believe that the designated respondents should give us better data, it would appear that the extra effort involved in reaching them does not yield a sufficiently large dividend.

About 7 percent of the workers whose earnings were reported on basis of some specific time unit (hourly, daily, weekly, etc.) responded, in answer to a further screening question, that they did receive some additional remuneration in the form of tips, commissions, bonuses, etc. The distribution of the group in terms of the amount they usually received was as follows:

Periodicity of receipt	Percent distribution by periodicity of receipt	Median receipts reported
Total	100.0	--
per day	13.4	$ 8
per week	20.0	23
per month	5.2	111
per year	49.3	458
other basis	2.1	152

While we have no idea as to how complete and accurate the above data might be, the test indicated that it is not impossible to collect some information about the additional earnings received by some workers in addition to their wages and salaries.

Future plans

Utilizing the results of this test and building upon the experience gained since 1967, the Bureau of Labor Statistics would like to initiate a more regular collection of data on weekly and hourly earnings through the CPS. One alternative being examined is monthly collection from the two outgoing rotation groups--the fourth and eighth month-in-sample groups. 6/ These limited monthly data, which would, in effect, be derived from one-fourth of the sample, could then be aggregated into reasonably reliable quarterly averages.

By using this option, there should be little if any impact on the other data derived from the CPS. But even this limited expansion would require some additional funds, and given the current budgetary situation it is not at all certain that they will be made available.

In any case, before launching any new major collection effort, the BLS, together with the Bureau of the Census, would like to conduct a validation test, where the reported earnings of a certain number of workers would be checked against the payroll records of their employers. Only then would we really know how accurately these earnings are reported.

FOOTNOTES

1/ The present size of the CPS sample, in terms of completed interviews, averages about 45,000 households a month distributed among 461 areas throughout the Nation. About 10,000 other households are being interviewed monthly in order to obtain better data for local areas, but the information from these additional households is not currently taken into account in the compilation of national averages.

2/ See Paul O. Flaim and Nicholas I. Peters, "Usual Weekly Earnings of American Workers," Monthly Labor Review, March 1972, pp. 28-38. and Thomas F. Bradshaw and John F. Stinson, "Trends in Weekly Earnings: An Analysis," Monthly Labor Review, August 1975, pp. 22-32.

3/ See Paul M. Ryscavage, "Measuring Union-Nonunion Earnings Differences," Monthly Labor Review, December 1974, pp. 3-9.

4/ An article on earnings foregone because of unemployment, authored by Paul M. Ryscavage and Curtis L. Gilroy, is scheduled to appear in an upcoming issue of the Monthly Labor Review.

5/ This procedure is similar to the one used in obtaining earnings data through the National Longitudinal Survey.

6/ Household falling in the CPS sample are visited (or called upon by telephone) for 4 consecutive months, are then dropped for 8 months, before being brought back into the sample for a second and final 4-month stint. Thus, one-eighth of the sample in any month consists of households in which interviews are being conducted for the fourth and final time of the initial 4-month stint, and another eighth consists of households who are in the last and final month of the second 4-month stint.

Table 1. Index of median usual weekly earnings of wage and salary workers by selected characteristics, in constant (1967) dollars, May 1967-May 1976

Characteristics	May 1967	May 1969	May 1970	May 1971	May 1972	May 1973	May 1974	May 1975	May 1976
ALL WAGE AND SALARY WORKERS									
Total	100.0	101.0	101.0	102.0	103.0	105.0	103.0	100.0	99.0
FULL-TIME WAGE AND SALARY WORKERS									
Total	100.0	101.0	101.8	103.6	104.5	110.0	105.5	105.5	105.5
Household status:									
Male head of household	100.0	102.3	103.8	106.8	106.1	114.4	111.4	109.8	109.8
Male relative of head	100.0	104.5	102.2	100.0	100.0	104.5	101.1	95.5	95.5
Male nonrelative of head	100.0	103.8	100.0	102.9	102.9	106.7	102.9	95.2	98.1
Female head of household	100.0	102.5	106.2	108.6	112.3	116.0	113.6	116.1	113.6
Wife of head	100.0	101.3	103.8	106.3	110.1	112.7	110.1	110.1	110.1
Female relative of head	100.0	100.0	101.4	102.8	104.2	104.2	100.0	100.0	100.0
Female nonrelative of head	100.0	111.6	111.6	114.5	121.7	121.7	114.5	126.1	123.2
Sex and age:									
Male, 16 years and over	100.0	103.2	104.0	106.3	107.1	113.5	111.1	110.3	109.5
16 to 24 years	100.0	101.0	99.0	95.9	96.9	105.1	102.0	95.9	95.9
25 years and over	100.0	103.0	104.5	107.6	108.3	116.7	113.6	112.1	112.1
Female, 16 years and over	100.0	101.3	103.8	106.4	109.0	112.8	109.0	110.3	110.3
16 to 24 years	100.0	101.4	102.7	101.4	104.1	105.4	102.7	98.6	100.0
25 years and over	100.0	102.5	105.1	107.6	111.4	116.5	113.9	116.5	115.2
Color:									
White	100.0	100.9	101.8	103.5	104.4	107.9	104.4	104.4	104.4
Male	100.0	102.3	103.8	106.1	105.3	112.2	109.9	107.6	107.6
Female	100.0	102.5	103.8	106.3	110.1	112.7	108.9	110.1	110.1
Negro and other races	100.0	105.1	108.9	112.7	116.5	124.1	121.5	124.1	121.5
Male	100.0	104.4	107.7	112.1	113.2	124.2	120.9	119.8	122.0
Female	100.0	106.3	111.1	114.3	125.4	128.6	127.0	130.2	130.2
Occupation:									
Professional and technical workers	100.0	104.8	106.8	106.8	105.5	110.3	107.5	105.5	103.4
Managers and administrators, except farm	100.0	98.8	99.4	100.6	104.2	109.7	104.2	104.2	103.6
Sales workers	100.0	99.1	100.9	102.6	106.1	108.8	103.5	104.4	102.6
Clerical workers	100.0	102.2	102.2	103.3	105.4	107.6	104.3	102.2	101.1
Craft and kindred workers	100.0	101.5	103.0	104.5	104.5	112.1	109.8	106.1	106.8
Operatives, except transport 1/	--	--	--	--	100.0	105.3	102.1	104.2	101.1
Transport equipment operatives 1/	--	--	--	--	100.0	105.7	101.6	101.6	103.3
Nonfarm laborers	100.0	103.2	101.1	103.2	104.3	111.7	108.5	103.2	101.1
Private household workers	100.0	96.9	103.1	96.9	100.0	93.8	106.3	106.3	109.4
Other service workers	100.0	100.0	100.0	105.3	110.7	112.0	106.7	102.7	105.3
Farm workers	100.0	105.2	105.2	105.2	110.3	125.9	125.9	120.7	122.4

1/ Separate data for these two groups not available prior to 1972.

Table 2. Distribution of earnings of workers paid at hourly rate, May 1973-May 1976
(Numbers in thousands)

Year	Total reporting earnings	Under $1.60	$1.60 to $1.99	$2.00 to $2.24	$2.25 to $2.49	$2.50 to $2.99	$3.00 to $3.99	$4.00 to $4.99	$5.00 to $5.99	$6.00 to $7.99	$8.00 to $9.99	$10.00 and over	Median	Mean
Workers in thousands														
1973	32,192	1,597	4,791	3,741	2,290	3,967	6,510	4,505	2,791	1,325	543	131	2.96	3.29
1974	32,152	1,108	2,284	4,982	2,159	3,984	6,656	4,655	3,496	2,032	610	186	3.20	3.54
1975	32,046	896	782	4,873	2,338	4,125	6,466	4,519	3,837	3,165	684	363	3.39	3.81
1976	34,237	758	482	2,457	3,995	4,763	7,406	4,564	3,975	4,355	914	569	3.55	4.06
Percent distribution														
1973	100.0	5.0	14.9	11.6	7.1	12.3	20.2	14.0	8.7	4.1	1.7	.4	--	--
1974	100.0	3.4	7.1	15.5	6.7	12.4	20.7	14.5	10.9	6.3	1.9	.6	--	--
1975	100.0	2.8	2.4	15.2	7.3	12.9	20.2	14.1	12.0	9.9	2.1	1.1	--	--
1976	100.0	2.2	1.4	7.2	11.7	13.9	21.6	13.3	11.6	12.7	2.7	1.7	--	--

Source: Current Population Survey

Table 3. Distribution of usual hourly earnings of all wage and salary workers, May 1973-May 1976
(Numbers in thousands)

Year	Total reporting earnings	Under $1.60	$1.60 to $1.99	$2.00 to $2.24	$2.25 to $2.49	$2.50 to $2.99	$3.00 to $3.99	$4.00 to $4.99	$5.00 to $5.99	$6.00 to $7.99	$8.00 to $9.99	$10.00 and over	Median	Mean
Workers in thousands														
1973	61,706	4,127	5,420	4,749	3,011	7,226	12,825	8,647	6,843	5,162	1,831	1,863	3.46	4.04
1974	61,220	2,805	3,043	5,752	2,867	6,823	12,611	8,796	7,557	6,530	2,163	2,274	3.71	4.34
1975	62,000	2,156	1,679	5,444	2,870	6,607	12,060	8,761	8,149	8,404	2,864	3,004	4.02	4.69
1976	63,010	1,644	1,183	3,103	3,863	6,835	12,590	8,634	8,212	9,679	3,367	3,901	4.26	4.99
Percent distribution														
1973	100.0	6.7	8.8	7.7	4.9	11.7	20.8	14.0	11.1	8.4	3.0	3.0	--	--
1974	100.0	4.6	5.0	9.4	4.7	11.1	20.6	14.4	12.3	10.7	3.5	3.7	--	--
1975	100.0	3.5	2.7	8.8	4.6	10.7	19.5	14.1	13.1	13.6	4.6	4.8	--	--
1976	100.0	2.6	1.9	4.9	6.1	10.8	20.0	13.7	13.0	15.4	5.3	6.2	--	--

Source: Current Population Survey

Table 4. Comparison of CPS earnings for workers paid at hourly rates with average hourly earnings from establishment survey, May 1975

Industry	CPS mean earnings (1)	Establishment mean earnings (2)	1/2 (3)	Establishment mean earnings excluding overtime (4)	1/4
Manufacturing	$4.22	$4.75	.89	$4.61	.92
Durable goods	4.49	5.06	.89	4.93	.91
Ordnance and accessories	6.20	5.15	1.20	4.99	1.24
Lumber and wood products	3.87	4.17	.93	4.02	.96
Furniture and fixtures	3.37	3.70	.91	3.64	.93
Stone, clay, and glass	4.33	4.83	.90	4.63	.94
Primary metals	4.93	6.04	.82	5.86	.84
Fabricated metal products	4.40	4.98	.88	4.85	.91
Machinery, except electrical	4.57	5.29	.86	5.12	.89
Electrical equipment	4.17	4.53	.92	4.45	.94
Transportation equipment	5.19	5.88	.88	5.73	.91
Instrument and related products	3.98	4.52	.88	4.43	.90
Miscellaneous manufacturing	3.53	3.75	.94	3.68	.96
Nondurable goods	3.80	4.30	.88	4.17	.91
Food and kindred products	3.96	4.52	.88	4.33	.91
Tobacco manufacturers	4.70	4.77	.99	4.69	1.00
Textile mill products	3.07	3.33	.92	3.22	.95
Apparel and other textile products	2.65	3.15	.84	3.11	.85
Paper and allied products	4.37	4.86	.90	4.66	.94
Printing and publishing	4.12	5.32	.77	N.A.	N.A.
Chemicals and allied products	4.61	5.30	.87	5.15	.90
Petroleum and coal products	5.65	6.33	.89	6.10	.93
Rubber and plastic products	3.97	4.30	.92	4.17	.95
Leather and leather products	2.78	3.20	.87	3.14	.89

MISSING DATA PROCEDURES: A COMPARATIVE STUDY

Barry L. Ford, Statistical Reporting Service, USDA

PURPOSE

The purpose of this paper is to discuss an investigation of the missing data problem in a list frame survey which has a simple stratified design. A description of the missing data procedures is given, and a general theoretical comparison is made among them. The main thrust of this report is a simulation experiment with these missing data procedures. All of the examples and data are from agricultural surveys by the Statistical Reporting Service (SRS) of the United States Department of Agriculture (USDA).

INTRODUCTION

In the area of survey design, the missing data problem is one of increasing concern. Nonresponse rates of 10% are not unusual for SRS surveys, and there is a fear that these rates may increase.

Why worry about missing data? If there is little difference between the missing data and the reported data in a simple stratified design, the only consequence of missing data is the reduction in sample size. This reduction can easily be offset by increasing the initial sample size. However, in many cases it is probable that the missing data and the reported data re not alike.

A difference between missing and reported data leads to biases in the survey estimates. The size of these biases depends on: 1. the magnitude of the difference between the missing and reported data. 2. the percentage of nonresponse.

Let: p = the percentage of the population which would respond

q = the percentage of the population which would not respond

μ = the population mean

μ_1 = the mean of the part of the population which would respond

μ_2 = the mean of the part of the population which would not respond

Of course, $\mu = p\mu_1 + q\mu_2$. Also, let

$$D = \mu_1 - \mu_2.$$

Then, the relative difference between the data which would be reported and the data which would not be reported is:

$$D' = \frac{\mu_1 - \mu_2}{\mu}.$$

The bias in only using the reported data to estimate μ is:

$$B = \mu_1 - \mu = q(\mu_1 - \mu_2) = qD.$$

Thus, the relationship between B, q and D is linear. Undoubtedly, these potential biases are the real cause of concern about missing data.

Similarly, the relationship between B' (the relative bias), q and D' is also linear:

$$B' = \frac{\mu_1 - \mu_2}{\mu} = q\frac{(\mu_1 - \mu_2)}{\mu} = qD'.$$

The causes of missing data are complex and varied, but the emphasis in any survey should be on eliminating or minimizing the likelihood of missing data before the survey starts. Procedures to estimate for missing data are a stopgap measure -- they are techniques to use after the survey is over when no other alternative is possible. Obviously, no procedure can be as good as not having any missing data. Furthermore, when the percentage of missing data is extremely high, there is probably no procedure that can estimate the missing data efficiently enough to make the survey worthwhile. With moderate and low missing data rates, perhaps some missing data procedures can minimize the bias to a tolerable level.

The six missing data procedures discussed in this investigation are the double sampling ratio procedure, the double sampling regression procedure, and four variations of the hot deck procedure. Some general advantages and disadvantages of each one are outlined.

The Double Sampling Ratio Procedure

Often there is an auxiliary variable associated with each sampling unit. This auxiliary variable may be a variable that is used to stratify the population, an observed variable, or any other additional variable that can be obtained for the whole sample. There should also be a reasonable correlation between the primary variable and the auxiliary variable.

The double sampling ratio design is well-known. In this experiment the first sample is the selected sample, including missing and reported data. Then the second sample is only the reported data. The ratio estimator and its approximate variance for a simple random sample (1, pg. 340) are:

$$\text{(I)}\quad \bar{y}_{Ratio} = \frac{\bar{y}}{\bar{x}}\,\bar{x}'$$

$$\text{(II)}\quad VAR(\bar{y}_{Ratio}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 - \left(\frac{1}{n} - \frac{1}{n'}\right)(2R\,S_yS_x - R^2S_x^2)$$

where:

$\bar{x}'$ = average of the auxiliary variable over the whole sample

$\bar{x}$ = average of the auxiliary variable over the part of the sample that reported data

$\bar{y}$ = average of the primary variable over the part of the sample that reported data

$\bar{X}$ = the average of the auxiliary variable over the whole population

$\bar{Y}$ = the average of the primary variable over the whole population

$R = \frac{\bar{Y}}{\bar{X}}$

S_x^2 = the variance of the auxiliary variable

S_y^2 = the variance of the primary variable

ρ = the correlation between x and y
n' = size of the entire sample
n = size of the sample that reported data
N = size of the population
(Note that the variance was multiplied by the finite population correction factor.)

Although the double sampling ratio estimator is almost always a biased estimator, it is easy to compute even for complex samples. In this report the design is a simple stratified sample so the above formulas are applied in each stratum. Usually S_y^2, S_x^2, ρ, and R are unknown, but their corresponding sample estimates can be substituted into the previous two equations (I and II). As Cochran points out (1, pg. 341), the resulting estimate of variance is not unbiased but appears to be a good approximation.

This ratio estimator makes two assumptions: 1. the initial sample is a random sample 2. the missing data comprise a *random* subsample of the initial sample. This second assumption is probably violated in most surveys; to what degree it is violated depends of course, on the particular situation. One hopes that the ratio estimate and its variance are fairly insensitive to a violation of the second assumption.

In essence the ratio estimator is a linear regression estimator with the intercept assumed to be zero. If the population does not follow the assumption of a linear model, then the ratio estimator (or any regression estimator) becomes a biased estimator. Researchers rarely accept the linear population model as completely realistic, but approximate analytical results and empirical studies show the bias is usually small (3, pg. 23-25; 7, pg. 208-209; 9).

One should remember that in a stratified design there also exists a combined ratio estimator. This estimator is used when the ratio $R = \frac{\bar{Y}}{\bar{X}}$ is equal in all strata. For the data in this study the idea that the ratios in all strata are equal is believed to be false. Thus, a separate ratio estimator is used in each stratum. However, the separate ratio estimator has an inherent danger of accumulating a serious bias across all strata. This accumulation is more likely to be serious when the stratum biases are in the same direction (1, pg. 168-173).

The Double Sampling Regression Procedure

The double sampling regression procedure is also quite common. Like the ratio procedure one has an auxiliary variable in addition to the primary variable. The formulas are (1, pg. 336-339):

(III) $\bar{y}_{Reg} = \bar{y} + b\,(\bar{x}' - \bar{x})$

(IV) $VAR\,(\bar{y}_{Reg}) = \frac{S_y^2\,(1-\rho^2)}{n} + \frac{\rho^2 S_y^2}{n'}$.

We will estimate $VAR\,(\bar{y}_{Reg})$ with:

$$var\,(\bar{y}_{Reg}) = \frac{s_{y.x}^2}{n} + \frac{s_y^2 - s_{y.x}^2}{n'}$$

Adjusting $var\,(\bar{y}_{Reg})$ by a finite population correction factor of $1\ \frac{n}{N}$, one obtains:

(V) $var'\,(\bar{y}_{Reg}) = (1 - \frac{n}{N})\left[\frac{s_{y.x}^2}{n} + \frac{s_y^2 - s_{y.x}^2}{n'}\right]$

as an estimate of the variance of $\bar{y}_{Reg}$ in a finite population where:

$\bar{x}'$, S_y^2, ρ^2, n', n, N are the same as for the ratio estimator

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$b = \frac{\sum_{i=1}^{n} (y_i - \bar{y})\,(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

$$s_{y.x}^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n} (y_i - \bar{y})^2 - b^2 \sum_{i=1}^{n} (x_i - \bar{x})^2\right].$$

As noted for the ratio estimator, the assumption of a linear population model is ignored. The estimators may then be biased, but one empirical study (3, pg. 22-25) lends support to the conjecture that these biases are small. Also, in the stratified data of this study a separate regression estimator is used as opposed to a combined regression estimator. Therefore, one again has the danger of accumulating a large bias across all strata, especially when each stratum bias is in the same direction (1, pg. 200-202).

Hot Deck Procedure

The hot deck is probably the most common missing data procedure in use at the present time, especially in complex surveys. The Bureau of the Census, the Statistical Reporting Service, Statistics Canada, and many others currently employ this missing data procedure. In spite of this wide use little testing or theoretical analysis on the impact of the hot deck procedure has ever been published (8). This situation is really not too surprising because, although the hot deck procedure is intuitively satisfying and extremely flexible; its flexibility and lack of a strong theoretical development deter anything but broad generalizations of its effects.

A basic outline of the hot deck procedure is:

1: Separate the sample into I classes based on k variables.
2: If an item is missing in a certain class, then randomly select a reported item from the same class.
3: Substitute the selected item for the missing item.
4: Compute sample estimates as if there are no missing values.

The most obvious consequence of this procedure one would challenge is the fact that the estimated variances of the sample mean are almost certainly biased below their actual values. Step

4 above allows one to use a sample size that includes the number of missing values. Thus, the loss of information due to missing data is not reflected in the sampling errors. For example, suppose two surveys cover the same population and have the same sample size. Furthermore, one survey has 30% missing data, and the other survey has no missing data. After applying the hot deck procedure, the errors of the estimates of these two surveys would probably be about equal. *Obviously, the standard errors from the survey that used the hot deck procedure should reflect the fact that 30% of the information is missing.*

One should also note that the *sample elements are no longer independent.* The hot deck procedure is essentially a duplicating process with reported values substituting for missing values. The covariance that results from this duplication is ignored in the hot deck procedure. Ignoring this covariance can be a serious error.

Probably the greatest attraction of the hot deck procedure is its operational simplicity. The classification of the data items into I classes is an extremely adaptable method. The classification variables may be cardinal, ordinal, categorical, etc. In fact, the whole classification method may vary from the subjective to the mathematically rigorous. In addition, many complex surveys will use the hot deck procedure because of the pressure to retain the planned sample design (eg. self-weighing designs, survey designs using balanced repeated replications, etc.). However, the looseness of the classification method has tended also to obstruct theoretical evaluations of the hot deck procedure and thus to impede any theoretical comparisons between it and other missing data procedures.

The hot deck does have some simple qualities to recommend it. For example, let $E(\bar{x}-\mu) = B$ be the bias associated with nonresponse when estimating the population mean, μ, with the mean of a simple random sample. To estimate μ using the hot deck procedure one divides the sample into I classes. Let $E(\bar{x}_i - \mu_i)$ be the bias in class i, i = 1, 2, ..., I. If p_i^* is the proportion of the population in class i, then the bias, B_{HD}, associated with the estimated mean, $\bar{x}_{HD}$, of the sample data after applying the hot deck procedure is simply:

$$B_{HD} = E(\bar{x}_{HD} - \mu) = \sum_{i=1}^{I} p_i^* B_i .$$

To prove this equation one notes:

$$E[\bar{x}_{HD}] = E\left[\sum_{i=1}^{I} p_i \bar{x}_i\right] = E_{n_i}\left[E\left\{\sum_{i=1}^{I} p_i \bar{x}_i \,\middle|\, n_i\right\}\right]$$

where n_i is the number of sample units that fell in class i, n is the total sample size, $p_i = \frac{n_i}{n}$, and $\bar{x}_i$ is the sample average for class i. The expected value inside the braces is over fixed n_i, and the expected value outside the braces is then over all possible values of n_i. Obviously,

$$E_{n_i}\left[E\left\{\sum_{i=1}^{I} p_i \bar{x}_i \,\middle|\, n_i\right\}\right] = E_{n_i}\left[\sum_{i=1}^{I} p_i \mu_i\right] = \sum_{i=1}^{I} p_i^* \mu_i .$$

In spite of the fact that $\bar{X}_i$ and n_i are not independent, they are uncorrelated. Now, if $|B_i| < |B|$, for each i then:

$$\text{(I)} \quad |B_{HD}| = \left|\sum_{i=1}^{I} p_i^* B_i\right| < \sum_{i=1}^{I} p_i^* |B_i| < \sum_{i=1}^{I} p_i^* |B| = |B|.$$

Thus, one can see that the bias using the hot deck procedure is less than the bias caused by omission data *on the condition that* $|B_i| < |B|$ for each i. This condition should hold in most cases, but there is no guarantee because it is a function of the quality of the classification method. A good classification method should decrease the absolute value of the bias below $|B|$ in each of the I classes. However, the hot deck allows any classification. The goodness of the classification process is left to the integrity of the statistician.

The "Closest" Procedure

One possible alternative to the random substitution of the hot deck procedure is to substitute the "closest" reported item for each of the missing items. With one auxiliary variable, the "closest" value to a missing item is simply the value for which the absolute difference between the auxiliary variable of the missing item and the auxiliary variable of the reported item is minimized. In the case of ties for the "closest" auxiliary value a random selection of one of the tied values is made.

This procedure should have the same effect as assigning the population to many strata and, selecting a few units from each stratum (since the stratification is based on the auxiliary variable). Thus, suppositions that the hot deck method improves with more narrowly defined strata can be examined with the results of the "closest" procedure.

Given a good range coverage, this procedure is fairly robust to very curved relationships between the auxiliary and primary variables. The data in this investigation is not curved enough to reveal this robust property of the "closest" procedure.

The "Two Closest" Procedure

This procedure is another variation of the hot deck procedure. Instead of substituting the "closest" reported item for each missing item, one substitutes the average of the "closest" value whose auxiliary value is smaller than the reported item and the "closest" value whose value is larger than the reported item.

The "Class" Mean Procedure

This last variation of the hot deck procedure substitutes the average of the reported units in a class for each missing unit in that class. It is the simplest and probably the cheapest of the procedures presented in this paper

THE SIMULATION EXPERIMENT

Why Use Simulation?

The need for simulation in this investigation is to compare the estimated variance of the estimated means. Possibly one might be able to compare how differences in the missing and reported data theoretically affect the estimated means using these six missing data procedures. However, the problem of analytically comparing the estimated variances of the estimated means is unreasonable. The fact that some assumptions fail in each procedure ties a knot in the analytical work.

For example, one should recall the estimated mean of the hot deck procedure, $\bar{x}_{HD}$. Assuming there are differences in the missing and reported data, one can not explicitly write the expected value of the estimated variance of $\bar{x}_{HD}$, $E[\hat{Var}(\bar{x}_{HD})]$. In fact, it is not known if $E[\hat{Var}(\bar{x}_{HD})] = Var(\bar{x}_{HD})$, and the author strongly doubts that it does. However, this paper will provide no evidence to support that supposition because the structure of the simulation of this experiment does not allow an estimate of $Var(\bar{x}_{HD})$ But does allow an estimate of $E[\hat{Var}(\bar{x}_{HD})]$. If $E[\hat{Var}(\bar{x}_{HD})] \neq Var(\bar{x}_{HD})$, then there is quite a weakness in the hot deck procedure. The costs of a simulation experiment providing this type of evidence would be much greater than the simulation actually used. This investigation contents itself with comparing the estimated variance of the estimated mean for each procedure with the estimated variance if the sample had no missing data. These comparisons will serve the purpose of revealing certain key qualities of each procedure.

One should note that the double sampling ratio and regression procedures also have variance estimates that involve assumptions and approximations that may be tenuous. For example, the assumption of a linear model is usually invalid in the regression and ratio procedures, and the ratio procedure simply uses substitution as an approximation to variance estimation. On the basis of two important studies (3;9) and practical experience one does not expect these biases in the variance estimates to be substantial for large samples. However, the comparisons among the procedures may be sensitive enough that these biases would be large enough to affect the comparisons.

Analysis

The primary point in the comparison of these procedures will be the minimization of the biases caused by missing data in the estimated means. Secondary importance is given to the comparisons of the estimated variances of the estimated means.

An important aspect of this study is the fact that the comparisons are based *on an experimental design where each observation is a result of a simulation of a procedure.* This situation is quite different from a simulation study where there may be a thousand or more simulations in order to narrow the confidence interval of an estimate almost to a point. Requests should be sent to the author for full details of the experimental design and details of the data used.

The correlations of the auxiliary variable and primary variable in the data range from 0.0 to 0.43. One may think that with larger correlations between the primary and auxiliary variable the estimates from the regression procedure would improve dramatically compared to the other procedures. However, the data in this study prevent evidence for or against this hypothesis. Surely, larger correlations would improve estimates resulting from all the procedures. Whether this improvement is equal for all the procedures is the question which can not be answered in this report.

Results

From the analysis of variance the six missing data procedures do not yield significantly different estimates of the mean. The average improvement in the estimated mean using the six missing data procedures is shown in Table 1. The relative bias reduction ranges from 8% to 26%.

Table 1: Average improvement in the estimated mean from the simulation of six missing data procedures.

B=Bias	A=Average estimate of mean Minus the true sample mean	$\frac{B-A}{B} \cdot 100\%$
-1.9	-1.40	26%
-2.9	-2.66	8%
-4.0	-3.63	9%
-6.2	-5.51	11%
-9.0	-7.65	15%
-14.0	-11.27	20%

Since there is no significant difference in the estimated means among the six procedures, the focus of interest becomes the estimated variances of the estimated means. The analysis of variance for the six missing data procedures with the estimated variance as the dependent variable was performed. Obviously, there is a significant difference among the estimated variances because the test statistic is so large:

$$\frac{MS_T}{MS_{TxPlots[Within\ AxB]}} = \frac{162.245}{0.391} = 414.95 .$$

Performing Duncan's multiple comparison test at a 95% significance level separates the procedures into the following groups:

{ t_1: double sampling ratio procedure
  t_2: double sampling regression procedure }

{ t_3: hot deck procedure with random substitution
  t_4: "closest" procedure
  t_5: "two closest" procedure

$\{t_6$: "class" mean procedure.

Perhaps it is more revealing to examine the estimated variances within each A x B cell, i.e. at different levels of bias. Table 2 shows the estimated variance using each procedure minus 9.685, the estimated variance if the sample has no missing data. Table 2 is on the next page.

The usual criterion for judging the variances of the estimated means resulting from the missing data procedures is that the smallest variance is the best. However, a procedure may result in a small estimated variance simply because of a large negative bias. By comparing the estimated variances resulting from the procedures with 9.685, the estimated variance if the sample has no missing data; one can judge if there are any large negative biases in the estimated variances. For example, if the estimated variance resulting from a missing data procedure is 7.20; then obviously, there is a large negative bias in estimating the variance.

One first notices that in Table 2, as in all the results, there is little difference between the regression and ratio estimates (procedures 1 and 2). One then notes that in the first three cases there is zero bias, and the hot deck estimator with random substitution (procedure 3) yields variances close to 9.685. The ratio and regression procedures have larger differences because they depend on the weak correlations of the primary and auxiliary variable. The "closest" procedure, the "two closest" procedure, and the "class" mean procedure (4, 5 and 6) have negative values. The negative values indicate *that their estimated variances are even less than the estimated variance if the sample has no missing data.*

CONCLUSIONS

The most important aspect in comparing these missing data procedures is to protect against biases in the estimated means (or totals). A split plot analysis of variance shows no significant differences among the estimated means which result in using these procedures. All the procedures reduce the relative bias that results from accepting the mean of the reported data as an estimate of the population mean. This reduction in relative bias is studied considering various non-response rates and considering various differences in the respondents and non-respondents. Varying from 8% to 26%, the reduction in relative bias averaged 15%. Considering the low correlations between the auxiliary and primary variables, this reduction is reasonable.

An important, though secondary, importance is attached to the estimated variances of the estimated means. All of these estimated variances except those from the ratio and regression procedures are generally less than the estimated variance that result with no missing data in the sample. Furthermore, this discrepancy *increases* as the relative bias increases. This part of the investigation clearly reveals why the hot deck, the "closest", the "two closest", and the "class" mean procedures may be undesirable. One is in the peculiar circumstance that, although all the procedures perform equally in reducing the bias of the mean, one can not choose a procedure on the basis of the smallest variance. The variances are deceiving. For example, using the hot deck procedure with the "class" mean substitution yields the smallest estimated variance, but this procedure is probably the worst. The results of this report indicate a large negative bias in the estimated variance resulting from the hot deck procedure because this estimated variance is much less than the sample variance when there is no missing data.

One should remember, however, that in many cases the only additional information on missing data is of a non-numerical nature. For example, data may be missing for a certain firm, but the only additional information is that it is a small insurance firm in Richmond, Virginia. In these cases, the hot deck procedure represents the only missing data procedure available. In the agricultural data of this investigation the additional information *is* numerical with the result that the ratio or regression procedure can be used and is better than the other procedures.

The final result of this investigation is a recommendation of the ratio or regression procedures (the effects of these two procedures being indistinquishable). These two procedures have been more theoretically explored than the other procedures. This estimated variances of the estimated means from the ratio or regression procedure reflect better than the other procedures the true quality of the data. The costs involved with the ratio regression computer program averaged about $5.00 simulation while the program for the other procedures averaged about $30.00 per simulation. Thus the computer costs of implementing the ratio or regression procedure are probably much lower.

BIBLIOGRAPHY

1. Cochran, William G., Sampling Techniques, John Wiley and Sons, Inc. 1963.
2. Dear, R.E., "A Principle Component Missing Data Method for Multiple Regression Models" Systems Development Corporation, Santa Monica, California, SP - 86, 1959.
3. Frankel, Martain R., Inference From Survey Samples: An Empirical Investigation. Institute for Social Research, University of Michigan, 1971.
4. Hartley, H.O. and Hocking, R.R., "The Analysis of Incomplete Data", Biometrics, Volume 27, pages 783-824, 1971.
5. Hocking, R.R. and Smith, W.B., "Estimation of Parameters in the Multiple Normal Distribution with Missing Observation", Journal of the American Statistical Association, Volume 63, pages 159-173, 1968.
6. Kirk, Roger E., Experimental Design: Procedures for the Behavioral Sciences, Brooks Cole Publishing Company, 1968.
7. Kish, Leslie, Survey Sampling, John Wiley and Sons, Inc. 1965.
8. Rockwell Richard C., "An Investigation of Imputation and Differential Quality of Data in the 1970 Census", Journal of the American Statistical Association, Volume 70, pg. 39-42, 1975.

Table 2 -- The difference between the estimated variance of the estimated mean resulting from a simulation of the missing data and 9.685, the estimated variance of the estimated mean if the sample has no missing data.

B=Bias	Estimated Variance of the Estimated Mean Minus 9.685					
	Double Sampling Ratio Procedure	Double Sampling Regression Procedure	Hot Deck Procedures			
			Random Substitution	"Closest"	"Two Closest"	"Class" Mean
0.0	0.556	0.556	0.126	-0.024	-0.045	-0.528
0.0	0.885	0.883	0.055	-0.701	-0.295	-1.238
0.0	2.478	2.478	0.580	-0.785	-0.415	-1.971
-1.9	0.263	0.263	-0.095	-0.416	-0.454	-0.781
-2.9	-0.223	-0.224	-0.654	-0.970	-0.906	-1.204
-4.0	0.355	0.353	-0.884	-0.973	-1.094	-1.685
-6.2	0.096	0.094	-0.950	-1.277	-1.385	-1.872
-9.0	1.087	1.084	-1.205	-1.434	-1.700	-2.928
-14.0	-0.086	-0.090	-1.928	-2.579	-2.750	-3.571
Average Over All Levels of Bias	0.601	0.600	-0.551	-1.018	-1.005	-1.753

FOR FULL DETAILS OF THIS RESEARCH ONE SHOULD WRITE TO BARRY L. FORD, ROOM 2920, SOUTH BUILDING, USDA, WASHINGTON, D.C., 20250

AN ASYMPTOTIC COVARIANCE STRUCTURE FOR TESTING HYPOTHESES ON RAKED CONTINGENCY TABLES FROM COMPLEX SAMPLE SURVEYS

Daniel H. Freeman, Jr., Yale University and Gary G. Koch, U. of North Carolina

0. Introduction

Recently, there have been substantial breakthroughs in the analysis of cross-classified frequency counts. One of the major algorithms for this type of analysis is variously known as iterative proportional fitting (IPF), raking, or the Deming and Stephan algorithm. The original version of the algorithm was discussed by Deming [1943]. Subsequently, there have been a number of papers and books on the techniques which utilize the algorithm (e.g. Bishop et al. [1975]). The aspect of the algorithm to be discussed in this paper is its application to complex survey data. It is assumed that such a random sample has a known probability structure.

Social and life scientists have been applying the algorithm to complex survey data for several years. Two examples are papers by Frederick J. Scheurin [1973] and Robert M. Hauser et al. [1975]. Scheurin states,

> "...when the sample size is large relative to the number of cells then substantively insignificant effects can become statistically significant. It also turns out to be quite difficult to make even approximate significance statements when data come from complex multistage samples..." (p.164).

This was stated in the context of a discussion concerning the use of log-linear models for purpose of generating hypotheses from a Current Population Survey data set. The focus was on poverty statistics. In the second example, Hauser et al. adjusted the "Occupational Changes in a Generation" data set downward by factor of 0.62 to reflect the efficiency of the survey design relative to simple random sampling (Hauser et al. [1975]: 282). However, they point out there should be additional adjustments for non-sampling error and simultaneous inference. Thus, the application of log-linear models to survey data requires dealing with two issues:

i. the relatively large size of the samples;
ii. the complexity of the survey design.

It is true that the size of the sample often leads to the statistical significance of largely uninterpretable interaction effects. However, this is not necessarily the case. For example, Freeman et al. [1977] discusses the fitting of relatively simple models to physician visit data from the National Health Interview Survey. This was a survey of about 40,000 households or 120,000 individuals. Moreover, the survey design was indirectly incorporated into that study. It is this issue which must be considered prior to deciding that the formal hypothesis testing as unnecessary because of the sample size.

The survey design is frequently too complex for the use of direct estimates of variance. However, for linear sample statistics techniques such as jack-knifing and pseudo-replication may be used to generate valid estimated covariance matrices. For "raked tables" the sample estimates are generally non-linear. Causey [1973] pointed out that Taylor series estimates are feasible for such tables. The tables are raked so as to minimize the "discrimination information." This is discussed for the simple survey situation in a number of places including Gokhale and Kullback [1976].

This paper shows that the problem is in fact a direct application of the "Functional Asymptotic Regression Methodology," Koch et al. [1975]. The key assumption is that the central covariance matrices are estimated either directly or indirectly by some method which accounts for the survey design. Previously, Koch et al. examined the problem where it could be assumed that the data were based on independent simple random samples. As noted in Freeman et al. [1977] the violation of this assumption in complex surveys can result is substantial reductions in the power of the test statistics. The discussion is in three parts. First, a general survey notation is presented. The raking model and its covariance matrix estimates are discussed. Lastly, an example is given.

1. Notation

Consider a set of d attributes. Let $j_g = 1,2,\ldots,L_g$ index the response categories for the g-th attribute where $g = 1,2,\ldots,d$. Let $\underset{\sim}{j} = (j_1,j_2,\ldots,j_d)$ denote the vector response profile. Let

$$N_{\underset{\sim}{j}\ell} = \begin{cases} 1 & \text{if element } \ell \text{ from population under study is classified as having response profile } \underset{\sim}{j}=(j_1,\ldots,j_d) \\ 0 & \text{if otherwise} \end{cases} \quad (1.1)$$

where $\ell = 1,2,\ldots,N$ with N being the total number of elements in the population. Let

$$U_\ell = \begin{cases} 1 & \text{if element } \ell \text{ from population is in sample} \\ 0 & \text{if otherwise} \end{cases} \quad (1.2)$$

The $\{U_\ell\}$ characterize the sample design including the nature of any clustering, stratification, and/or multistage selection. Let $\phi_\ell = E\{U_\ell\}$ denote probability of selection for element ℓ from population. Let $n = \sum_{\ell=1}^{N} \phi_\ell$ denote the sample size.

The multivariate relationships among the d attributes can be summarized in terms of the d-dimensional contingency table of weighted frequencies

$$\hat{N}_{\underset{\sim}{j}} = \hat{N}_{j_1 j_2 \cdots j_d} = \sum_{\ell=1}^{N} \frac{1}{\phi_\ell} U_\ell N_{j_1 j_2 \cdots j_d, \ell} . \quad (1.3)$$

In this framework, let $p_{\underset{\sim}{j}} = (\hat{N}_{\underset{\sim}{j}}/N)$ denote the corresponding relative frequency or weighted observed sample proportion. The $p_{\underset{\sim}{j}}$ are unbiased estimators for the parameters

$$\pi_{\underset{\sim}{j}} = \pi_{j_1 j_2 \cdots j_d} = \frac{1}{N} E\{\hat{N}_{j_1 j_2 \cdots j_d}\} = \frac{1}{N} \sum_{\ell=1}^{N} N_{j_1 j_2 \cdots j_d, \ell} \quad (1.4)$$

which reflect the average distribution of the respective response profiles in the population. Let $\hat{\underset{\sim}{N}}$, $\underset{\sim}{p}$, and $\underset{\sim}{\pi}$ denote the vectors

$$\hat{\underset{\sim}{N}} = \begin{bmatrix} \hat{N}_{11\ldots1} \\ \cdots \\ \hat{N}_{L_1 L_2 \ldots L_d} \end{bmatrix}, \quad \underset{\sim}{p} = \begin{bmatrix} p_{11\ldots1} \\ \cdots \\ p_{L_1 L_2 \ldots L_d} \end{bmatrix}$$

$$\underset{\sim}{\pi} = \begin{bmatrix} \pi_{11\ldots1} \\ \cdots \\ \pi_{L_1 L_2 \ldots L_d} \end{bmatrix} \quad (1.5)$$

2. Methods of Adjustment

A sample from a specific population may be regarded as yielding two types of information:

A. Estimators for marginal distributions of certain subsets of attributes. This type of information is called "allocation structure."

B. Estimators for higher order measures of association and/or interaction which reflect the relationships across the marginal subsets in (A). This type of information is called "association structure."

With these considerations in mind, the observed sample can be adjusted to provide estimators for other target populations of interest if the following assumptions hold:

i. the target population has KNOWN "allocation structure" via census or other sample survey data

ii. the target population has the SAME "association structure" as the sampled population.

Examples of such target populations include:

a. various local (county or state) subdivisions of a nationally sampled population

b. other local, national, or international target populations which may partially overlap a sampled local population.

More specifically, let $\underset{\sim}{\pi}$ denote the parameter vector which characterizes the distribution of the response profiles for the sampled population, and let $\underset{\sim}{p}$ denote its corresponding estimator. Let $\underset{\sim}{\pi}_T$ denote the parameter vector for the target population. Let $\underset{\sim}{A}_T$ denote a matrix of coefficients whose columns generate the pertinent marginal distributions comprising the known "allocation structure," and let $\underset{\sim}{\xi}_T$ denote their corresponding known values. Thus, assumption (i) means that $\underset{\sim}{\pi}_T$ satisfies

$$\underset{\sim}{A}_T' \underset{\sim}{\pi}_T = \underset{\sim}{\xi}_T \quad (2.1)$$

where without loss of generality, $\underset{\sim}{A}_T'$ will be regarded as having full rank by deletion of unnecessary rows. The matrix $\underset{\sim}{A}_T'$ also reflects the fact that the elements of $\underset{\sim}{\pi}_T$ satisfy the constraint

$$\underset{\sim}{1}' \underset{\sim}{\pi}_T = 1 \quad (2.2)$$

where $\underset{\sim}{1}'$ is a row vector of 1's. Given the formulation (2.1) of "allocation structure," attention will be directed at the asymptotic covariance structure of the estimator for the parameter vector $\underset{\sim}{\pi}_T$ which is obtained by applying assumption (ii) with respect to an appropriate definition of "association structure."

2.1. Adjustment with respect to complete "association structure"

Let $\underset{\sim}{K}$ denote an ortho-complement matrix to $\underset{\sim}{A}_T$. Then assumption (ii) means that $\underset{\sim}{\pi}_T$ satisfies

$$\underset{\sim}{K}'\{\log_e(\underset{\sim}{\pi}_T)\} = \underset{\sim}{K}'\{\log_e(\underset{\sim}{\pi})\} \quad (2.3)$$

where, in this context, "association structure" is formulated in terms of log-linear contrast functions. If $\underset{\sim}{p}$ denotes the sample estimator of $\underset{\sim}{\pi}$ defined by (1.5), then (2.1) and (2.3) imply that the marginal adjustment (raking) estimator $\hat{\underset{\sim}{\pi}}_T$ of $\underset{\sim}{\pi}_T$ is characterized by the equations

$$\underset{\sim}{A}_T' \hat{\underset{\sim}{\pi}}_T = \underset{\sim}{\xi}_T \quad (2.4)$$

$$K'\{\log_e(\hat{\pi}_T)\} = K'\{\log_e(p)\}. \tag{2.5}$$

Within this framework, the estimator $\hat{\pi}_T$ may be determined (provided both assumptions (i) and (ii) are true so that a solution π_T to (2.1) and (2.3) almost always exists if the sample size n is sufficiently large) by applying the Deming-Stephan Iterative Proportional Fitting (IPF) algorithm to adjust an initial estimator which satisfies (2.5) to conform successively with each of the respective marginal configurations which comprise the "allocation structure" equations (2.4) since such operations preserve the agreement of successive solutions with the "association structure" equations (2.5).

The asymptotic covariance matrix of the estimators $\hat{\pi}_T$ is obtained by the well-known δ-method (based on the first order Taylor series) with the required first derivative matrix being determined by implicit techniques. In this regard, if both sides of (2.4) - (2.5) are differentiated with respect to p, it follows that

$$\begin{bmatrix} A_T' \\ K'D_{\hat{\pi}_T}^{-1} \end{bmatrix} \begin{bmatrix} \dfrac{d\hat{\pi}_T}{dp} \end{bmatrix} = \begin{bmatrix} 0 \\ K'D_p^{-1} \end{bmatrix} \tag{2.6}$$

where D_y denotes a diagonal matrix with the elements of y on the diagonal. Since

$$\begin{bmatrix} A_T' \\ K'D_{\hat{\pi}_T}^{-1} \end{bmatrix}^{-1} = \{D_{\hat{\pi}_T} A_T [A_T' D_{\hat{\pi}_T} A_T]^{-1}, K[K'D_{\hat{\pi}_T} K]^{-1}\} \tag{2.7}$$

the equations (2.6) may be solved to yield

$$\left[\frac{d\hat{\pi}_T}{dp} \,\middle|\, p=\pi \right] = K[K'D_{\pi_T}^{-1}K]^{-1}K'D_{\pi}^{-1} \tag{2.8}$$

Thus, if $V(\pi)$ denotes the covariance matrix of the estimator p, then the asymptotic covariance matrix for $\hat{\pi}_T$ is given by

$$V_{\hat{\pi}_T}(\pi) = K[K'D_{\pi_T}^{-1}K]^{-1}K'D_{\pi}[V(\pi)]D_{\pi}^{-1}K[K'D_{\pi_T}^{-1}K]^{-1}K'. \tag{2.9}$$

For univariate problems, a similar result is found in Causey (1972).

If the sample design is simple random sampling (with replacement), then

$$V(\pi) = \frac{1}{n}\{D_{\pi} - \pi\pi'\}. \tag{2.10}$$

Thus, for this special case, the covariance matrix (2.9) may be simplified to

$$V_{\hat{\pi}_T}(\pi) = K[K'D_{\pi_T}^{-1}K]^{-1}K'D_{\pi}^{-1}K[K'D_{\pi_T}^{-1}K]^{-1}K'/n \tag{2.11}$$

Moreover, if the target population is identical to the sampled population (which is the case when the "allocation structure" of the population under study is known a priori as is the case with samples from registration systems like licensed drivers), then π_T and π are identical so that (2.11) may be further simplified to

$$V_{\hat{\pi}_T}(\pi) = K[K'D_{\pi}^{-1}K]^{-1} K'/n \tag{2.12}$$

An analogous simplification could also be applied to the more general result (2.9) for this situation.

Reasonable estimators for the covariance matrix $V_{\hat{\pi}_T}(\pi)$ may be constructed by replacing $V(\pi)$ by an appropriate consistent estimator V_p which is obtained by either direct or replication methods and replacing π and π_T by p and $\hat{\pi}_T$ respectively. The resulting estimated covariance matrix $V_{\hat{\pi}_T}$ may be used in conjunction with $\hat{\pi}_T$ to test various hypotheses by weighted least squares methods. In this regard, an appropriate test statistic for the hypothesis

$$H_0\colon\ C\pi_T = 0, \tag{2.13}$$

where C is assumed to be a full rank matrix, is the Wald statistic

$$Q_C = \hat{\pi}_T' C' [C V_{\hat{\pi}_T} C']^{-1} C \hat{\pi}_T \tag{2.14}$$

which has approximately a chi-square distribution with D.F. = Rank(C) in large samples.

2.2. Adjustment with respect to reduced "association structure"

For certain situations, it may be possible to assume that the vector π is characterized by a log-linear model

$$\pi = \pi(\beta) = \{\exp(X\beta)\}/\{1'[\exp(X\beta)]\}, \tag{2.15}$$

where X is a known full rank design matrix whose columns represent a basis for the main effects and interactions which constitute the model and β is an unknown parameter vector. When the model (2.15) holds, a reasonable estimator for π may be obtained by solving the equations

$$X'[\pi(\hat{\beta})] = X'p \ . \qquad (2.16)$$

If the matrix X has an hierarchical structure which includes with any given interaction variable all other interaction variables of the same type and all corresponding lower order interactions, the equations (2.16) may be solved by applying the Deming-Stephan IPF algorithm to adjust an initial estimator

$$\hat{\pi}_0 = \prod_{g=1}^{d} L_g^{-1}\{1\} \qquad (2.17)$$

which trivially satisfies the model (2.15) to conform successively with each of the respective marginal configurations which are associated with the equations (2.16).

The asymptotic covariance matrix of the estimator $\hat{\beta}$ of β is obtained by the δ-method with the first derivative matrix being determined by implicit techniques. In this regard, if both sides of (2.16) are differentiated with respect to p, it follows that

$$X' \frac{d}{dp} \frac{\exp(X\beta)}{1'\{\exp(X\beta)\}} = X'$$

$$X'[D_{\hat{\pi}} - \hat{\pi}\hat{\pi}']X \frac{d\beta}{dp} = X' \qquad (2.18)$$

where $\hat{\pi} = \{\exp(X\hat{\beta})/1'[\exp(X\hat{\beta})]\}$. The equations (2.18) may be solved to yield

$$\left.\frac{d\hat{\beta}}{dp}\right|_{p=\pi} = \{X'[D_\pi - \pi\pi']X\}^{-1}X' . \qquad (2.19)$$

Thus, the asymptotic covariance matrix $V_{\hat{\beta}}(\pi)$ for $\hat{\beta}$ is given by

$$V_{\hat{\beta}}(\pi) = \{X'[D_\pi - \pi\pi']X\}^{-1}X'V(\pi)X\{X'[D_\pi - \pi\pi']X\}^{-1} \qquad (2.20)$$

If the sample design is simple random sampling, then (2.10) may be used to simplify (2.20) to

$$V_{\hat{\beta}}(\pi) = \frac{1}{n}\{X'[D_\pi - \pi\pi']X\}^{-1} . \qquad (2.21)$$

Moreover, it can be noted that $\hat{\beta}$ represents the maximum likelihood estimator of β for this situation.

A reasonable estimator for $V_\beta(\pi)$ may be constructed by replacing $V(\pi)$ by an appropriate estimator V_p as described previously and replacing π by $\hat{\pi}$. The resulting estimated covariance matrix V_β together with $\hat{\beta}$ provide a framework for further analysis by weighted least squares methods.

The results (2.15) - (2.21) may be applied to the marginal adjustment (raking) situation by noting that (2.15) implies that the "association structure" equations (2.3) may be written as

$$\begin{bmatrix} K' \\ X_C' \end{bmatrix} \{\log_e(\pi_T)\} = \begin{bmatrix} K' \\ X_C' \end{bmatrix} \{\log_e(\pi)\} = \begin{bmatrix} K'X\beta \\ 0 \end{bmatrix} \qquad (2.22)$$

where X_C is an ortho-complement matrix to $[A_T, X]$ and K is an ortho-complement matrix to $[A_T, X_C]$. By proceeding as described in (2.4) - (2.9), it can be verified that the asymptotic covariance matrix for $\hat{\pi}_T$ under the model (2.15) is given by

$$V_{\pi_T}(\beta) = H[V_\beta(\pi)]H' \qquad (2.23)$$

where

$$H = [K, X_C] \begin{bmatrix} K'D_{\pi_T}^{-1}K , & K'D_{\pi_T}^{-1}X_C \\ X'D_{\pi_T}^{-1}K , & X_C'D_{\pi_T}^{-1}X_C \end{bmatrix}^{-1} \begin{bmatrix} K'X \\ 0 \end{bmatrix}$$

Further extensions of these results may be undertaken by allowing ξ_T to be replaced by an estimator $\hat{\xi}_T$ which is either independent or correlated with p.

3. Example

The data in Table 1 have been used by Ireland and Kullback [1968; pp. 707-713] to illustrate the application of IPF for the adjustment of a contingency table to a known marginal "allocation structure." They are being reanalyzed here to indicate the reduction in variance which is achieved by using such "raking" procedures to estimate the cell probabilities π.

These data originally come from a study undertaken by Roberts et al. [1939; Biometrika, 31, pp. 56-66]. The experimental design involves $n = 3734$ mice from a single population, each of which is classified with respect to the presence or absence of the attributes A, B, and D. The "allocation structure" of interest is defined in terms of the hypothesis that the probability of the presence (or absence) of each separate attribute is (1/2). Thus, with respect to the matrix notation in (2.1), it follows that

$$\underset{\sim}{A}_T' = \begin{bmatrix} 1&1&1&1&1&1&1&1 \\ 1&1&1&1&0&0&0&0 \\ 1&1&0&0&1&1&0&0 \\ 1&0&1&0&1&0&1&0 \end{bmatrix}, \quad \underset{\sim}{\xi}_T = \begin{bmatrix} 1.0 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} \qquad (3.1)$$

The "association structure" which is to be preserved in the sense of (2.3) corresponds to the log-linear functions

$$\underset{\sim}{F}(\underset{\sim}{\pi}) = \underset{\sim}{K}'[\log_e(\underset{\sim}{\pi})] \qquad (3.2)$$

where

$$\underset{\sim}{K}' = \begin{bmatrix} 1&1&-1&-1&-1&-1&1&1 \\ 1&-1&1&-1&-1&1&-1&1 \\ 1&-1&-1&1&1&-1&-1&1 \\ 1&-1&-1&1&-1&1&1&-1 \end{bmatrix}, \qquad (3.3)$$

which pertain to the first and second order interactions among the three attributes.

By using IPF to adjust the observed frequencies in Table 1 to the "allocation structure" specified by (3.1), Ireland and Kullback obtain the "raked" predicted cell frequencies shown in Table 2. The corresponding predicted proportions $\hat{\underset{\sim}{\pi}}_T$ are also given there together with their respective standard errors based on (2.12). Thus, by comparing these results with their counterparts in Table 1, it can be noted that the predicted proportions $\hat{\underset{\sim}{\pi}}_T$ are very similar to the original observed proportions, but have substantially smaller estimated standard errors.

Finally, since the "allocation structure" (3.1) corresponds to an hypothesis rather than a priori known constraints, Ireland and Kullback indicate that its acceptability for these data is supported by a nonsignificant ($\alpha = .25$) Minimum Discrimination Information Chi-Square Statistic for goodness of fit $Q_{MDI}(\hat{\underset{\sim}{\pi}}_T|\underset{\sim}{p}) = 3.42$ with D.F. = 3.

ACKNOWLEDGMENTS

This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-05 and the GM-00047 and by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 75-2. The authors would like to thank Stephen E. Fienberg and Frederick J. Scheurin for several comments which were helpful during the initial stages of this research. Additional insights were gained as a part of general discussions of statistical methodology during the Social Science Research Council 1975 Summer Workshop on Crime and Criminal Justice Statistics. Finally, we would like to thank Alcinda Lewis and Jane Beth Markley for computational assistance, and Joann Raccio, Karen McKee, and Rebecca Wesson for their cheeerful and conscientious typing of the manuscript.

REFERENCES

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis, MIT Press, Cambridge.

Causey, B.D. (1972). Sensitivity of raked contingency table totals to changes in problem conditions, Annals of Mathematical Statistics 43: 656-658.

Deming, W.E. (1943). Chapter VII: Adjusting sample frequencies to expected marginal totals, in Statistical Adjustment of Data, Dover Publications, Inc., New York: 96-127.

Freeman, D.H., Freeman, J.L., Brock, D.B., and Koch, G.G. (1977). Strategies in the multivariate analysis of data from complex surveys II: An application to the U.S. National Health Interview Survey, International Statistical Review, accepted for publication.

Gokhale, D.V. and Kullback, S. (1976). An information analysis of multinomial experiments and contingency tables: The general linear hypothesis, presentation to Washington Statistical Society.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). Analysis of categorical data by linear models, Biometrics 25: 489-504.

Hauser, R.M., Koffel, J.N., Travis, H.P., and Dickinson (1975). Temporal change in occupational mobility: Evidence for men in the United States, American Sociological Review 40: 279-297.

Ireland, C.T. and Kullback, S. (1968). Minimum discrimination information estimation, Biometrics 24: 707-713.

Koch, G.G., Freeman, D.H., and Freeman, J.L. (1975), Strategies in the multivariate analysis of data from complex surveys, International Statistical Review 43: 59-78.

Koch, G.G., Freeman, D.H., and Tolley, H.D. (1975). The asymptotic covariance structure of estimated parameters from contingency table log-linear models, Institute of Statistics Mimeo Series No. 1046, University of North Carolina.

Scheurin, F.L. (1973). Ransacking CPS tabulations: applications of the log-linear model to poverty statistics, Annals of Economic and Social Measurement 2: 159-182.

1. TABULATION OF MICE ACCORDING TO ATTRIBUTES A, B, AND D

	Response profile for attributes A vs B vs D							
Attribute A	Y	Y	Y	Y	N	N	N	N
Attribute B	Y	Y	N	N	Y	Y	N	N
Attribute D	Y	N	Y	N	Y	N	Y	N
Overall group observed cell frequency	475	460	462	509	467	440	494	427
Observed proportions	0.1272	0.1232	0.1237	0.1363	0.1251	0.1178	0.1323	0.1144
Estimated s.e.	0.0055	0.0054	0.0054	0.0056	0.0054	0.0053	0.0055	0.0052

Y denotes presence of the attribute; N denotes absence.

2. "RAKED" PREDICTED CONTINGENCY TABLE FOR ATTRIBUTES A, B, AND D

	Response profile for attributes A vs B vs D							
Attribute A	Y	Y	Y	Y	N	N	N	N
Attribute B	Y	Y	N	N	Y	Y	N	N
Attribute D	Y	N	Y	N	Y	N	Y	N
Overall group "raked" predicted cell frequency	463.3	464.5	438.7	500.5	475.4	463.8	489.6	438.2
Predicted Proportions	0.1241	0.1244	0.1175	0.1340	0.1273	0.1242	0.1311	0.1174
Estimated s.e.	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041	0.0041

Y denotes presence of the attribute; N denotes absence.

SAMPLE-DEPENDENT ESTIMATION FOR THE NATIONAL SURVEY OF FAMILY GROWTH

Dwight K. French and E. Earl Bryant, National Center for Health Statistics

Introduction

The first cycle of the National Survey of Family Growth (NSFG) was conducted by the National Center for Health Statistics under a contractual arrangement with the National Opinion Research Center (NORC) of the University of Chicago between July 1973 and February 1974. The survey was based on a rather complex multi-stage area probability sample of about 10,000 women 15-44 years of age, who had ever been married, or who had never been married but had natural-born children currently living in the household. Personal interviews were conducted to collect information about fertility, family planning and maternal and child health practices. The purpose of this paper is to present the methodology and results of a study to determine the best estimator for the survey based on sample results, a technique known as sample-dependent estimation.

The choice of estimators for the NSFG was between a simple inflation-type estimator based on reciprocals of selection probabilities and a post-stratified estimator using variable bench marks.

Post-stratified estimators are in common use for national surveys which cover the general population, since census-type information needed for bench marks is readily available. For surveys such as the Health Examination Survey and the Current Population Survey, it is known that post-stratification by age, race and sex produces more precise results than are attainable by a simple inflation estimator.[1]

Unfortunately, current census counts for the NSFG target population were not available for the time period that the survey was conducted. It was possible, however, to obtain estimates based on the Bureau of the Census's Current Population Survey (CPS), which collects information on a sample of about 50,000 households each month. Based on our knowledge of the efficiency of post-stratified estimators using census bench marks, it seemed reasonable that a large part of the gain over simple inflation estimators could be achieved using estimates of the bench marks if the estimates were precise enough. Furthermore, the use of large samples reduces the risk of making a wrong decision among alternative sample-dependent estimators.[2]

The criterion for choosing between the two estimators was based on the magnitude of their respective sampling variances. Values of over 200 statistics were calculated using both estimators, and variances were computed using identical balanced half-sample replication procedures. If poststratification decreased the variances of many estimates and did not cause large increases in variance for more than a few, the poststratified estimator would be selected; otherwise, the simple inflation estimator would be used.

Alternative NSFG Estimators

Let $Y = Y_1 + Y_2$ represent an aggregate parameter for ever-married women 15-44 years of age (Y_1) and never-married women in the same age group with natural-born children present in the home at the time of the interview (Y_2). Alternative estimators of Y considered for the NSFG were

$$Y'_{NP} = Y'_1 + Y'_2 \quad \text{and} \tag{1}$$

$$Y'_P = Y''_1 + Y'_2 \quad \text{where} \tag{2}$$

Y'_1 and Y'_2 are simple inflation estimators and

$$Y''_1 = \sum_{\alpha=1}^{12} Y'_{\alpha 1} \frac{X^*_{\alpha 1}}{X'_{\alpha 1}} \quad \text{is a poststratified} \tag{3}$$

estimator.

The $X'_{\alpha 1}$ represent NSFG estimates of ever-married women in each of 12 age-race classes, and the $X^*_{\alpha 1}$ are the corresponding CPS estimates.

Computation of CPS Estimates of Ever-Married Women

Since the NSFG was conducted between July 1973 and February 1974, population values were needed for September 1973, the approximate mid-point of the survey. One possibility was to adjust the March 1973 CPS estimates of ever-married women. Although the total CPS sample is very large, the number of women in some of the desired age-race groups, especially for black women, would be relatively small. Since the CPS sample design allows an accumulation of the sample over time, it is possible to improve the precision of estimates by combining several CPS monthly samples. For the NSFG, two estimates of ever-married women were computed for each of 12 age-color groups using CPS data collected in March of 1970, 1971, 1972, and 1973. One estimator was the average of the four CPS values; the other was a predicted value based on simple linear regression. The predicted value was used if the regression coefficient was significantly different from zero. Otherwise, the average was used. The estimates and their approximate relative standard errors are shown in Table 1.

Variance Computation for the Two Estimators

Variances for Y'_P and Y'_{NP} were based on the same set of 48 balanced half-sample replicates. Y'_{kNP}, the nonpoststratified estimate for the k^{th} half-sample, was calculated in exactly the same way as Y'_{NP}, except that all records were given an additional weight of 2 to compensate for the selection of the half-sample. Y'_{kP}, the corresponding poststratified half-sample estimate, was calculated in the same way as Y'_{kNP} except that only records for single women were given the additional weight. The poststratification process inflated the estimate for ever-married women to its proper size. Given estimates Y'_{kNP} and Y'_{kP} from each half-sample,

Table 1. Population Estimates Based on CPS Data, and Their Relative Standard Errors

Age	Black: Population	Black: Rel. Std. Error (%)	Other: Population	Other: Rel. Std. Error (%)
15-19	115,000	5.2	993,000	1.9
20-24	564,000	2.7	4,967,000	0.8
25-29	622,000	2.4	6,153,000*	1.9
30-34	602,000	2.5	5,371,000*	1.8
35-39	575,000	2.6	4,751,000	0.8
40-44	593,000	2.5	4,940,000*	0.6

*predicted values

the variance Y'_P was estimated by

$$s^2_{Y'_P} = \frac{1}{48} \sum_{k=1}^{48} (Y'_{kP} - Y'_P)^2 \quad (4)$$

and the variance of Y'_{NP} was estimated by

$$s^2_{Y'_{NP}} = \frac{1}{48} \sum_{k=1}^{48} (Y'_{kNP} - Y'_{NP})^2. \quad (5)$$

Additional Variance of Y'_P Due to Variance in the $X^*_{\alpha 1}$

The half-sample replication procedure described in the previous section omitted one obvious source of variability in the poststratified estimator. Because each of the 48 half-sample estimates of characteristic Y for ever-married women in the 12 age-color classes was poststratified to the same set of CPS bench marks, the procedure implicitly assumed that the CPS values were known constants. In fact, these values are estimates, subject to sampling variability.

Because the NSFG was designed to interview only one woman from each sample household, estimates for ever-married and never-married women are essentially uncorrelated, as are estimates for women in different age-race classes; i.e.,

$$\text{Cov}\,(Y'_1, Y'_2) = 0$$

and

$$\text{Cov}\,(Y'_{\alpha 1} \frac{X^*_{\alpha 1}}{X'_{\alpha 1}}, Y'_{\beta 1} \frac{X^*_{\beta 1}}{X'_{\beta 1}}) = 0, \ \alpha \neq \beta$$

Therefore, the variance of Y'_P can be expressed as

$$\sigma^2_{Y'_P} = \sum_{\alpha=1}^{12} \text{Var}\,(Y'_{\alpha 1} \frac{X^*_{\alpha 1}}{X'_{\alpha 1}}) + \text{Var}\,(\sum_{\alpha=1}^{12} Y'_{\alpha 2}) \quad (6)$$

The second term of (6) is properly estimated by the half-sample procedure, and will be omitted from the remainder of the discussion.

The approximate sampling variance of

$$Y''_{\alpha 1} = \frac{Y'_{\alpha 1}}{X'_{\alpha 1}} X^*_{\alpha 1},$$

the estimator for an individual α class, is[3]

$$\sigma^2_{Y'_{\alpha 1}} = \left(\frac{Y_{\alpha 1}}{X_{\alpha 1}}\right)^2 (X_{\alpha 1})^2 \left[\frac{\text{Var}\,(Y'_{\alpha 1}/X'_{\alpha 1})}{(Y_{\alpha 1}/X_{\alpha 1})^2} + \frac{\text{Var}\,(X^*_{\alpha 1})}{(X_{\alpha 1})^2} + \frac{2\,\text{Cov}\,\{(Y'_{\alpha 1}/X'_{\alpha 1}),(X^*_{\alpha 1})\}}{(Y_{\alpha 1}/X_{\alpha 1})(X_{\alpha 1})}\right] \quad (7)$$

The last term inside the brackets is zero since the NSFG and the CPS are independent surveys.

Substituting $Y'_{\alpha 1}$, $X'_{\alpha 1}$, and $X^*_{\alpha 1}$ into (7) yields the estimate:

$$\hat{\sigma}^2_{Y''_{\alpha 1}} = \left(\frac{Y'_{\alpha 1}}{X'_{\alpha 1}}\right)^2 (X^*_1) \left[\frac{\hat{\text{Var}}\,(Y'_{\alpha 1}/X'_{\alpha 1})}{(Y'_{\alpha 1}/X'_{\alpha 1})^2} + \frac{\text{Var}\,(X^*_{\alpha 1})}{(X^*_{\alpha 1})^2}\right]$$

or

$$\hat{\sigma}^2_{Y''_{\alpha 1}} = (X^*_{\alpha 1})^2\, \hat{\text{Var}}\,(Y'_{\alpha 1}/X'_{\alpha 1}) + \left(\frac{Y'_{\alpha 1}}{X'_{\alpha 1}}\right)^2 \sigma^2_{X^*_{\alpha 1}} \quad (8)$$

The first term of (8) has been computed by the replication procedure; the second term is the component that must be added on. The computation for estimates of Y for individual α classes is straightforward, since $Y'_{\alpha 1}$, $X'_{\alpha 1}$ and $\sigma^2_{X^*_{\alpha 1}}$ are all known for each α class.

For estimates of Y in any population group G covering more than one α class (such as national estimates, estimates for all eligible black females, or estimates for all eligible females age 20-24) application of (6) and (8) yields

$$\sigma^2_{Y'_P} = \sum_{\alpha \varepsilon G} (X^*_{\alpha 1})^2\, \sigma^2_{(Y'_{\alpha 1}/X'_{\alpha 1})} + \sum_{\alpha \varepsilon G} \left(\frac{Y'_{\alpha 1}}{X'_{\alpha 1}}\right)^2 \sigma^2_{X^*_{\alpha 1}} \quad (9)$$

Again, the first summation in (9) is measured by the half-sample variance estimator, while the

Table 2. Poststratified and Nonpoststratified Estimates of the Number of Currently Married U.S. Women, Their Relative Standard Errors, and the Percent Reduction in RSE due to the Poststratified Estimator, by Wife's Religion and Age, National Survey of Family Growth, 1973.

Wife's Religion and Age	Y'_{NP} (1)	RSE of Y'_{NP} (2)	Y'_P (3)	RSE of Y'_P Excluding CPS Component (4)	RSE of Y'_P Including CPS Component (5)	Percent Reduction in RSE: $\frac{(2)-(5)}{(2)}$ (6)
Catholic						
All Ages	7,382,079	.0609	7,661,129	.0575	.0578	5.09
15-19	264,354	.1462	269,556	.1114	.1128	22.85
20-24	1,218,068	.0819	1,319,352	.0763	.0768	6.23
25-29	1,595,694	.0736	1,737,304	.0606	.0636	13.59
30-34	1,599,122	.0928	1,585,113	.0788	.0804	13.36
35-39	1,465,395	.0887	1,393,582	.0814	.0817	7.89
40-44	1,239,446	.0837	1,356,222	.0798	.0801	4.30
Jewish						
All Ages	433,130	.1146	449,013	.1145	.1147	-0.09
15-19	8,391	.7093	8,695	.7173	.7175	-1.16
20-24	30,868	.3721	33,499	.3796	.3797	-2.04
25-29	102,139	.1869	111,566	.1820	.1830	2.09
30-34	74,593	.2434	73,887	.2396	.2401	1.36
35-39	112,247	.2362	106,722	.2193	.2195	7.07
40-44	104,872	.2321	114,644	.2341	.2342	-0.90
Protestant						
All Ages	16,759,541	.0367	17,297,274	.0241	.0248	32.43
15-19	667,533	.0993	662,219	.0607	.0633	36.25
20-24	3,050,430	.0643	3,255,852	.0347	.0357	44.48
25-29	3,586,495	.0467	3,872,671	.0277	.0337	27.84
30-34	3,391,181	.0491	3,349,635	.0345	.0381	22.40
35-39	3,191,781	.0519	3,017,746	.0366	.0373	28.13
40-44	2,872,121	.0492	3,139,151	.0299	.0305	38.01
Other-None						
All Ages	1,188,020	.0822	1,238,770	.0634	.0637	22.51
15-19	85,805	.1996	87,255	.2067	.2075	-3.96
20-24	318,845	.1145	340,689	.1132	.1135	0.87
25-29	314,563	.1292	341,180	.1117	.1133	12.31
30-34	241,614	.1843	239,690	.1685	.1693	8.14
35-39	120,273	.1530	113,626	.1481	.1483	3.07
40-44	106,920	.1739	116,330	.1822	.1823	-4.83

second summation is not. A problem for these estimates is that the $Y'_{\alpha 1}$ are not readily available. One method of approximating the second sum is to assume that $Y_{\alpha 1} = C \cdot X_{\alpha 1}$ for all α. Under this assumption, a natural estimate of $Y'_{\alpha 1}$ is

$$Y'_{\alpha 1} = \frac{Y'_P \cdot X'_{\alpha 1}}{\sum_{\alpha \varepsilon G} X'_{\alpha 1}} \qquad (10)$$

and the extra component of variance is given by

$$\sum_{\alpha \varepsilon G} \left(\frac{Y'_{\alpha 1}}{X_{\alpha 1}}\right)^2 \sigma^2_{X^*_{\alpha 1}} = \sum_{\alpha \varepsilon G} \left[\frac{(Y'_P)^2 \ (X'_{\alpha 1})^2}{(\sum_\alpha X'_{\alpha 1})^2}\right] \cdot \left[\frac{\sigma^2_{X^*_{\alpha 1}}}{(X'_{\alpha 1})^2}\right]$$

$$= \sum_{\alpha \varepsilon G} \left[\frac{(Y'_P)^2}{(\sum_\alpha X'_{\alpha 1})^2}\right] \cdot \sigma^2_{X^*_{\alpha 1}}$$

$$= \frac{(Y'_P)^2}{(\sum_\alpha X'_{\alpha 1})^2} \sum_{\alpha \varepsilon G} \sigma^2_{X^*_{\alpha 1}} \qquad (11)$$

which can be computed.

Results

Simple inflation and poststratified estimates of the number of currently married U.S. women and the relative standard errors (RSE) of both estimates were computed for more than 200 subdomains of the population. Tabulations were made by race, wife's education, husband's education, family income, wife's religion, ethnic origin, labor force status and working status, each cross-classified by the wife's age. Table 2 shows the statistics produced for one cross-classification, namely wife's religion and wife's age. The poststratified estimator was better than the inflation estimator for more than 80 percent of the domains. The improvement in precision was 20 percent or better for almost two-fifths of the domains (Table 3).

The percent reductions in RSE for the tabulations by wife's education, income, religion and working status were each tested by the Wilcoxon Signed-Rank Test. The null hypothesis for each table was that the two estimators were equally precise; i.e., the average percent reduction in RSE due to the poststratified estimator was 0. The null hypothesis was rejected at the two-tailed .01 level of significance for all four tables. The remaining four tables generally showed similar results, which was expected because wife's education and husband's education are correlated, labor force status and working status are highly correlated, and race and ethnic origin age groups closely correspond to the poststratification classes. The one exception to the pattern occurred for persons of Spanish origin. The RSE of the poststratified estimator was higher for five of the six age groups, but never more than 6 percent higher.

As a result of these findings, the poststratified estimator was adopted for the NSFG.

References

1. Simmons, Walt R. and Bean, Judy A., Impact of Design and Estimation Components on Inference, *New Developments in Survey Sampling,* Wiley--InterScience, New York, 1969.

2. Dalenius, Tore and Ziegler, Walter Joh. (Ed.), Sample Dependent Estimation in Survey Sampling. *Contributions to Applied Statistics dedicated to Arthur Linder,* Birkhauser Verlag, 1976.

3. Hansen, Morris H., Hurwitz, William N. and Madow, William G., *Sample Survey Methods and Theory,* Vol. 1, John Wiley & Sons, N.Y., 1953, pg. 513.

Table 3. Distribution of the percent reduction in relative standard error of the simple inflation estimator by use of poststratified estimator. National Survey of Family Growth, 1973.

		Percent Reduction*						
	Total	-10 to 0	0-9	10-19	20-29	30-39	40-49	50+
Number of estimates	215	35	55	43	32	18	13	19
Percent of estimates	100.0	16.3	25.6	20.0	14.9	8.4	6.0	8.8

$$\text{*Percent Reduction} = \frac{\text{RSE } (Y'_{NP}) - \text{RSE } (Y'_P)}{\text{RSE } (Y'_{NP})}$$

PROPERTIES OF THE CORRECTION FOR ATTENUATION ESTIMATOR

Wayne A. Fuller, Iowa State University
Michael A. Hidiroglou, Statistics Canada

ABSTRACT

The limiting distribution of the regression coefficients calculated from a correlation matrix that has been corrected for attenuation is obtained. Methods of estimating the covariance matrix of the vector of regression coefficients are presented. Nonnormal regression variables and nondiagonal error matrices are considered. The procedures are illustrated with data on the socioeconomic career.

1. INTRODUCTION

The effect of measurement error upon estimated regression coefficients has long been recognized. Cochran [5], Johnston [9], Walker and Lev [16] and Wiley [18] are recent references reporting the distortions that are introduced into standard regression statistics when the independent variables are measured with error. In a regression with a single independent variable the regression coefficients, on average, are reduced in absolute value, attenuated, when compared to those computed in the absence of measurement error. The same is true of the correlation coefficients.

In some areas it is possible to obtain good estimates of the ratio of the measurement error variance to the total variance. If the measurement errors in different independent variables are uncorrelated, the estimated variance ratios can be used to adjust the observed correlation matrix to construct an estimate of the correlation matrix one would obtain in the absence of measurement errors. The resulting estimated correlation matrix is said to have been corrected for attenuation. Regression equations can then be estimated from the correlation (or covariance) matrix corrected for attenuation. Although the method has been extensively used in the social sciences, little discussion of the sampling properties of the estimators is available (see Bohrnstedt and Carter [3]).

In this paper we derive the limiting distribution for the correction for attenuation estimator for both the uncorrelated and correlated measurement error cases. We also demonstrate how the standard error of the regression coefficients can be estimated when the error and (or) the true values have an arbitrary distribution with finite fourth moments.

The distributional results are illustrated using the causal chain model for the socioeconomic career discussed by Featherman [6] and Kelley [12].

2. MODEL AND ESTIMATION

We write the model as

$$\underset{\sim}{Y} = \underset{\sim}{x}\,\underset{\sim}{\beta} + \underset{\sim}{e}$$
$$\underset{\sim}{X} = \underset{\sim}{x} + \underset{\sim}{u}\ , \qquad (2.1)$$

where $\underset{\sim}{Y}$ is an $n \times 1$ vector, $\underset{\sim}{x}$ is an $n \times k$ matrix, and $\underset{\sim}{\beta}$ is a $k \times 1$ vector. The vector $\underset{\sim}{Y}$ and the matrix $\underset{\sim}{X}$ are observed and an estimator of $\underset{\sim}{\beta}$ is desired. The matrix $\underset{\sim}{u}$ is the matrix of measurement errors. We shall utilize the following assumptions:

(i) The vectors of errors $(e_t, \underset{\sim}{u}_t)$, t=1, 2,..., where $\underset{\sim}{u}_t$ is the t^{th} row of $\underset{\sim}{u}$ are distributed as normal independent random variables with zero mean and covariance matrix

$$\begin{pmatrix} \sigma_e^2 & \underset{\sim}{\Sigma}_{eu} \\ \underset{\sim}{\Sigma}_{ue} & \underset{\sim}{\Sigma}_{uu} \end{pmatrix} = \text{diag}(\sigma_e^2, \sigma_{u_1}^2, \sigma_{u_2}^2, \ldots, \sigma_{u_k}^2)\ .$$

(ii) The distribution of $(e_j, \underset{\sim}{u}_j)$ is independent of that of $\underset{\sim}{x}_t$ for all t, j where $\underset{\sim}{x}_t$ is the t^{th} row of $\underset{\sim}{x}$.

(iii) The $\underset{\sim}{x}_t$, t = 1,2,...,n, are distributed as normal independent random variables with mean $\underset{\sim}{0}$ and nonsingular covariance matrix $\underset{\sim}{\Sigma}_{xx}$.

The reader will note that we have lost no generality in assuming the mean of the $\underset{\sim}{x}_t$ to be zero. If the mean is unknown we make an orthogonal transformation to reduce the problem to the stated form. In practice one uses the corrected sums of squares and products in the analysis when the mean in unknown. If an independent variable is measured without error, then $\sigma_{u_i}^2 = 0$ for that variable.

Since $\underset{\sim}{x}_t$ and $\underset{\sim}{u}_t$ are normally distributed it follows that $\underset{\sim}{X}_t = \underset{\sim}{x}_t + \underset{\sim}{u}_t$, t = 1,2,...,n are distributed as normal independent random variables with mean zero and nonsingular covariance matrix $\underset{\sim}{\Sigma}_{XX} = \underset{\sim}{\Sigma}_{xx} + \underset{\sim}{\Sigma}_{uu}$. It is also assumed that:

(iv) The ratios λ_i, i = 1,2,...,k, of error variance $\sigma_{u_i}^2$ to total variance $\sigma_{X_i}^2$, where $\sigma_{u_i}^2$ is the i^{th} diagonal element of $\underset{\sim}{\Sigma}_{uu}$ and $\sigma_{X_i}^2$ is the diagonal element of $\underset{\sim}{\Sigma}_{XX}$, are known.

The quantity $(1-\lambda_i)$ is called the reliability of the i^{th} variable. We denote the diagonal matrix of ratios by

$$\underset{\sim}{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)\ . \qquad (2.2)$$

We define the correction for attenuation estimator of $\underset{\sim}{\beta}$ by

$$\hat{\underset{\sim}{\beta}} = \hat{\underset{\sim}{H}}^{-1}(n^{-1}\,\underset{\sim}{X}'\,\underset{\sim}{Y})\ , \qquad (2.3)$$

where

$$\hat{H} = \begin{cases} \frac{1}{n} X'X - D\Lambda D & , \text{ if } \hat{f} \geq 1 + n^{-1} \\ \frac{1}{n} X'X - (f - n^{-1}) D\Lambda D , & \text{ if } \hat{f} < 1 + n^{-1} , \end{cases}$$

$$D = \text{diag}(s_{X_1}, s_{X_2}, \ldots, s_{X_k}) ,$$

$$s^2_{X_i} = n^{-1} \sum_{t=1}^{n} X^2_{ti} ,$$

$\hat{f}$ is the smallest root of

$$|M - fTGT| = 0 ,$$

$$M = \frac{1}{n} \begin{pmatrix} Y'Y & Y'X \\ X'Y & X'X \end{pmatrix} ,$$

$$T = \begin{bmatrix} s_Y & 0 \\ 0 & D \end{bmatrix} ,$$

$$s^2_Y = n^{-1} \sum_{t=1}^{n} Y^2_t ,$$

$$G = \begin{cases} \text{diag}(0, \lambda_1, \lambda_2, \ldots, \lambda_k) & \text{if the reliability of Y is unknown} \\ \text{diag}(\lambda_{ee}, \lambda_1, \lambda_2, \ldots, \lambda_k) & \text{if the reliability of Y is known and denoted by } \lambda_{ee} . \end{cases}$$

The slight modification introduced by the calculation of $\hat{f}$ guarantees that the matrix $\hat{H}$ to be inverted is always positive definite, and that the estimated covariance matrix of the true variables is positive definite. In practice, if one obtains a small $\hat{f}$ one should investigate the hypothesis that the covariance matrix Σ_{xx} is singular by computing the smallest root of

$$|n^{-1} X'X - \ell D\Lambda D| = 0 .$$

If $\hat{\ell}$ is not significantly different from one, it may be desirable to modify the model by reducing the dimension of X . By the results of Fuller [7], the distribution of $(n-k)\hat{f}$ is approximately that of a chi-square random variable with n-k degrees of freedom when the rank of $(x:y)'(x:y)$, where $y = x\beta$, is k and the reliability of Y is known. Similarly, $(n-k+1)\hat{\ell}$ is approximately distributed as a chi-square random variable with n-k+1 degrees of freedom when the rank of $x'x$ is k-1 .

Theorem 1: Let model (2.1) and assumptions (i) through (iv) hold. Then

$$n^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, \Sigma_{xx}^{-1} C \Sigma_{xx}^{-1}) ,$$

where the ijth element of the matrix C is

$$c_{ij} = \sigma_{X_iX_j}\left(\sigma^2_v - 2\lambda^2_i\beta^2_i\sigma^2_{X_i} - 2\lambda^2_j\beta^2_j\sigma^2_{X_j} + 2\lambda_i\lambda_j\beta_i\beta_j\sigma_{X_iX_j}\right) + \lambda_i\lambda_j\beta_i\beta_j\sigma^2_{X_i}\sigma^2_{X_j}$$

and $\sigma^2_v = \sigma^2_e + \sum_{i=1}^{k} \beta^2_i\sigma^2_{u_i}$.

Proofs of the theorems may be obtained by writing the authors for the complete manuscript.

The covariance matrix of $\hat{\beta}$ is estimated by replacing the parameters by their estimates, where σ^2_v is estimated by

$$s^2_v = \frac{1}{n-k} \sum_{t=1}^{n} (Y_t - X_t\beta)^2 ,$$

$\hat{H}$ is an estimator of Σ_{xx}, $n^{-1} X'X$ furnishes estimators of $\sigma_{X_iX_j}$, and X_t is the tth row of the matrix X .

In the computations sums of squares corrected for the mean will typically be used throughout. If an intercept term is computed for the regression,

$$\tilde{\beta}_0 = \bar{Y} - \bar{X}'\tilde{\beta} ,$$

the variance of the estimated intercept can be estimated by

$$\hat{V}\{\tilde{\beta}_0\} = n^{-1} s^2_v + \bar{X}'\hat{H}^{-1} \hat{C}\hat{H}^{-1} \bar{X} ,$$

where $\bar{X}' = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k)$ and $\hat{C}$ is the estimator of C .

The form of the covariance matrix obtained in Theorem 1 was a function of the moment properties of the normal distribution. However, the fact that the estimator converged in distribution to a normal random variable required only independence of the observations and the existence of certain moments. Therefore, we can extend the procedure to nonnormal distributions. We also relax the assumption that the covariance matrix of the measurement errors is diagonal. We make the assumptions:

(v) The vectors (e_t, u_t, x_t), t = 1,2,..., are independently and identically distributed with

$$E\{e_t, u_t\} = 0$$

$$E\{x_t\} = \mu$$

$$E\{(e_t, u_t)'(e_t, u_t)\} = \Sigma,$$

$$E\{(x_t - \mu)'(x_t - \mu)\} = \Sigma_{xx}$$

$$E\{x_t' (e_t, u_t)\} = 0,$$

and finite fourth moments, where Σ_{xx} is nonsingular.

(vi) The matrices Λ_{eu} and Λ_{uu} are known, where

$$G^{\dagger} = \mathcal{D}^{-1} \Sigma \mathcal{D}^{-1} = \begin{pmatrix} \lambda_{ee} & \Lambda_{eu} \\ \Lambda_{ue} & \Lambda_{uu} \end{pmatrix} ,$$

$$\mathcal{D} = \text{diag}(\sigma_Y, \sigma_{X_1}, \sigma_{X_2}, \ldots, \sigma_{X_k}), \text{ and}$$

$$\Sigma = \begin{pmatrix} \sigma_e^2 & \Sigma_{eu} \\ \Sigma_{ue} & \Sigma_{uu} \end{pmatrix} .$$

The estimator analogous to that defined in (2.3) is

$$\tilde{\beta} = \hat{H}^{-1}(n^{-1} X'Y - D \Lambda_{ue} s_Y) , \qquad (2.5)$$

where $\hat{H}$ and D are defined below equation (2.3) with Λ_{uu} replacing Λ and $G^{\dagger}$ replacing G. If λ_{ee} is unknown it is set equal to $\Lambda_{eu} \Lambda_{uu}^{-1} \Lambda_{ue}$ in the calculation of $\hat{f}$.

Theorem 2: Let model (2.1) with assumptions (v) and (vi) hold. Then

$$n^{\frac{1}{2}} \Sigma_{xx}(\tilde{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, A) ,$$

where

$$A = E\{d_t' d_t\} ,$$

$$d_t = (d_{t1}, d_{t2}, \ldots, d_{tk}) ,$$

$$d_{ti} = X_{ti}v_t - \frac{1}{2}\left[\lambda_{u_ie}\left(\frac{\sigma_Y}{\sigma_{X_i}} X_{ti}^2 + \frac{\sigma_{X_i}}{\sigma_Y} Y_t^2\right) - \sum_{j=1}^{k} \lambda_{u_iu_j} \beta_j\left(\frac{\sigma_{X_i}}{\sigma_{X_j}} X_{tj}^2 + \frac{\sigma_{X_j}}{\sigma_{X_i}} X_{ti}^2\right)\right]$$

X_{ti} is the ti^{th} element of X, v_t is the t^{th} element of v, and λ_{u_ie} is the i^{th} element of Λ_{ue}.

The form of the result presented in Theorem 2 suggests an estimator of the variance of $\tilde{\beta}$ that is relatively easy to compute.

Theorem 3: Let model (2.1) with assumptions (v) and (vi) hold. Then $\hat{H}^{-1} A \hat{H}^{-1}$, where

$$\tilde{A} = (n-k)^{-1} \sum_{t=1}^{n} \tilde{d}_t' \tilde{d}_t ,$$

$$\tilde{d}_t = (\tilde{d}_{t1}, \tilde{d}_{t2}, \ldots, \tilde{d}_{tk}),$$

$$\tilde{d}_{ti} = X_{ti} \hat{v}_t - \frac{1}{2}\left[\lambda_{u_ie}\left(\frac{s_Y}{s_{X_i}} X_{ti}^2\right) + \frac{s_{X_i}}{s_Y} Y_t^2 - \sum_{j=1}^{k} \lambda_{u_iu_j} \tilde{\beta}_j\left(\frac{s_{X_i}}{s_{X_j}} X_{tj}^2 + \frac{s_{X_j}}{s_{X_i}} X_{ti}^2\right)\right] ,$$

$$\hat{v}_t = Y_t - \sum_{j=1}^{k} \tilde{\beta}_j X_{tj} ,$$

is a consistent estimator of the covariance matrix of the limiting distribution of $n^{\frac{1}{2}}(\tilde{\beta} - \beta)$.

3. EXAMPLE

To illustrate the computations associated with the correction for attenuation, we use some data studied by Featherman [6] and Kelley [12]. (See also the Comments section of The American Sociological Review (1973, p. 785-796.) The data were kindly made available by Professor Featherman. The data pertain to the careers of 715 white native American urban married males. The reader is referred to the cited articles for a complete description of the data. Two of the several equations estimated in the original studies are:

$$Q_3 = \beta_1 Q_F + \beta_2 T + \beta_3 Q_1 + \beta_4 Q_2$$

$$Q_2 = \alpha_1 Q_F + \alpha_2 T + \alpha_3 Q_1 + \alpha_4 I_1$$

where

- Q_i, i = 1,2,3 is occupation at time i, where i = 1 is at marriage, time 2 is about eight years after marriage and time 3 is about sixteen years after marriage.
- Q_F is father's occupation
- T is years of formal education
- I_1 is income in thousands of dollars at time one.

Occupation is recorded on an eleven point scale based upon the 1947 National Opinion Research Center study [13].

Kelley gave the reliabilities for the variables as 0.718 for father's occupation, 0.933 for education, 0.861 for occupation, and 0.852 for income.

Considerable interest centered on the coefficients β_3 and α_4. Under one theoretical model both of these coefficients were hypothesized to be zero. Estimates of the two equations are given below.

$$Q_3 = \underset{\substack{(0.040)\\(0.043)}}{0.094Q_F} + \underset{\substack{(0.035)\\(0.034)}}{0.137T} - \underset{\substack{(0.074)\\(0.078)}}{0.044Q_1} + \underset{\substack{(0.085)\\(0.091)}}{0.661Q_2}$$

$$Q_2 = \underset{\substack{(0.036)\\(0.034)}}{0.080Q_F} + \underset{\substack{(0.038)\\(0.034)}}{0.176T} + \underset{\substack{(0.040)\\(0.054)}}{0.651Q_1} - \underset{\substack{(0.030)\\(0.026)}}{0.097I_1}$$

The first set of numbers in parentheses are the estimated standard errors computed under the assumption of normality. The second set are the estimated standard errors computed under the more general assumptions. The coefficients are reported in the original units. Also the method used to treat missing values differed from that used by Featherman. Therefore, the coefficients are not identical to those reported by Featherman and Kelley. From a substantive viewpoint the coefficient for Q_1 in equation one could easily be zero. However, if one accepts the assumptions it is very unlikely that the coefficient for income in the second equation is zero.

The variables are clearly not normal because all are restricted to a few integer values. Procedures based on normality gave estimated standard errors very similar to those obtained under the more general assumptions for the first equation. On the other hand, the estimated standard error for Q_1 in the second equation computed under the normal assumption is quite different from that computed under the more general assumptions.

The joint distribution of Q_2 and Q_1 deviates considerably from normality. For example, the residuals from the ordinary regression of Q_2 on Q_1, say $\hat{\delta}$, have a coefficient of skewness of 0.34 and a kurtosis of 3.32. The approximate standard errors of these quantities, under normality, are 0.09 and 0.18, respectively. There is also considerable evidence that the conditional mean of Q_2 given Q_1 is not linear, the t statistic for the quadratic term in a regression of Q_2 on Q_1 and Q_1^2 being 6.7. Also the conditional variance of Q_2 given Q_1 is not constant, the regression of the squared regression residuals, $\hat{\delta}^2$, on Q_1 and Q_1^2 give an F-statistic of 30.8 with two and 712 degrees of freedom.

Because of the robustness of the general procedure it is recommended unless the sample size is very small.

ACKNOWLEDGEMENT

This research was partly supported by the Bureau of the Census through Joint Statistical Agreements J.S.A. 75-1, and J.S.A. 76-66.

REFERENCES

[1] Blalock, H. M., Jr., "Estimating Measurement Error Using Multiple Indicators and Several Points in Time," American Sociological Review, 35 (February 1969), 101-11.

[2] Bock, R. D. and Peterson, A. C., "A Multivariate Correction for Attenuation," Biometrika, 62 (December 1975), 673-78.

[3] Bohrnstedt, G. W. and Carter, T. M., "Robustness in Regression Analysis," in H. L. Costner, ed., Sociological Methodology, 1971, San Francisco: Jossey-Bass, Inc., 1971, 118-46.

[4] Booth, G. D., "The Errors-in-Variables Model When the Covariance Matrix is Not Constant," Ph.D. dissertation, Iowa State University, 1973.

[5] Cochran, W. G., "Errors of Measurement in Statistics," Technometrics, 10 (November 1968), 637-66.

[6] Featherman, D. L., "A Research Note: A Social Structural Model for the Socioeconomic Career," Amer. J. Soc., 77 (1971), 293-304.

[7] Fuller, W. A., "Properties of Estimators in the Errors-in-Variables Model," paper presented at the 1971 annual meeting of the Econometric Society, 1971 (mimeographed).

[8] Hidiroglou, M. A., "Estimation of Regression Parameters for Finite Populations," Ph.D. dissertation, Iowa State University, 1974.

[9] Johnston, J., Econometric Methods, New York: McGraw-Hill Book Co., 1963.

[10] Jöreskog, K. G., "A General Method for Analysis of Covariance Structures," Biometrika, 57 (August 1970), 239-51.

[11] Jöreskog, K. G., "Statistical Analysis of Sets of Congeneric Tests," Psychometrika, (June 1971), 109-33.

[12] Kelley, J., "Causal Chain Models for the Socioeconomic Career," American Sociological Review, 38 (1973), 481-93.

[13] National Opinion Research Center, "Jobs and Occupations: A Popular Evaluation," Opinion News, 9 (September 1947), 3-13.

[14] Malinvaud, E., Statistical Methods of Econometrics, Amsterdam: North-Holland Publishing Co., 1966.

[15] Rao, C. R., Linear Statistical Inference and Its Applications, New York: Wiley, 1965.

[16] Walker, H. M. and Lev, J., Statistical Inference, New York: Holt, Rhinehart & Winston, Inc., 1953.

[17] Warren, R. D., Keller, J. K., and Fuller, W. A., "An Errors-in-Variables Analysis of Managerial Role Performance," J. Amer. Statist. Assoc., 69 (December 1974), 886-93.

[18] Wiley, D. E., "The Identification Problem for Structural Equations Models with Unmeasured Variables," in A. S. Goldberger and O. D. Duncan, eds., Structural Equation Models in the Social Sciences, New York: Seminar Press, 1973.

New Projections of the U.S. Labor Force to 1990

Howard N. Fullerton, Jr. and Paul O. Flaim, Bureau of Labor Statistics

During the latter half of the seventies, the civilian labor force of the United States is estimated to expand by 9.1 million, reaching 101.7 million by 1980. This increase, which will complete the absorption of the post World War II "baby boom" into the labor force, implies an average annual growth rate of 1.9 percent, somewhat below the growth rate of the early seventies. After 1980, the rate of growth should decrease even further, averaging only 1.1 percent a year during the eighties. Even at this lower rate, however, the labor force is expected to rise to 108.6 million in 1985 and to 113.8 in 1990.

The labor force participation rates of women have been increasing rapidly for many years, and this increase is projected to continue, although at a lesser pace, through 1990 (table 1 and 2). The number of women in the labor force, 37.0 million in 1975, is projected to rise by 11.6 million by 1990, an increase of 1.8 percent per year. This represents a slightly slower rate of growth than experienced over the early seventies, reflecting a projected increase in births during the late seventies and early eighties and, most of all, the drop after 1985 in the number of persons at the ages at which people enter the labor force.

Although the labor force participation rate of men is expected to continue its gradual historical decline over the same period, the male civilian labor force, which numbered 55.6 million in 1975, is still expected to grow by 9.6 million between 1975 and 1990, or by 1.1 percent a year (table 1). Despite the slower rate of growth for men than for women, men are still expected to comprise 57 percent of the labor force in 1990.

In terms of its age composition, civilian labor force growth between 1975 and 1990 will be heavily concentrated in the central age groups. As shown below, the youth labor force (age 16-24) is actually expected to decline in size by 1990, and the elderly labor force (age 55 and over) is projected to be only slightly higher in 1990 than in 1975.

Age group	1975 actual (000)	1980	1985	1990
16-24 years	22,266	24,266	23,067	20,952
25-54 years	56,182	62,497	70,616	78,567
55 years and over	14,165	14,910	14,919	14,320

These estimates of future labor force growth, which are discussed in detail below, were derived as part of the Bureau of Labor Statistics' periodic reassessment of its projections of the growth trends for the various sections of the American economy. 1/ As described in detail at the end of this report, these labor force projections were made essentially by extrapolating the observed trends in the labor force participation rates of the various groups of working age and by then applying the results to the Bureau of the Census projected population levels for these groups. These population estimate are not subject to much uncertainty, as they deal with persons who have already been born and who, by and large, can be counted. There is, however, a much greater element of uncertainty with regard to the future course of the labor force participation rates, of the various population groups, and it is within this context that these labor force projections should be viewed.

Changes in the youth labor force

The principal reason for the projected slowdown in labor force growth during the 1980's is the sharp drop in the birth rate during the 1960's which will result in a smaller number of youth reaching labor force age than in recent years. In fact the youth labor force (age 16 to 24) is already growing at a slower pace. From 22.3 million in 1975, it is projected to increase by 2.0 million to 1980, a much smaller change than the 4.4 million advance between 1970 and 1975 (table 2). After 1980, the youth labor force is actually expected to decline, dropping by 1.2 million by 1985 and by a further 2.1 million, to 20.9 million, by 1990 (table 1). The shrinking of the youth labor force, will occur first among teenagers and then among youths 20 to 24.

Growth of the teenage labor force is actually projected to level off during the late 1970's after increasing only slightly, by about 330,000, from its 1975 level. Over the eighties, however, their labor force is expected to decrease by 1.5 million, reflecting the sharp drop in the teenage population. Because of the drop in their population, it is expected that by 1990 teenagers will represent only 6.7 percent of the labor force, a significantly smaller proportion than the 9.5 percent which they accounted for in 1975.

With large numbers of the present teenage cohorts passing through their early 20's over the next few years, the labor force group age 20 to 24 should grow very rapidly until the early 1980's, at which point it, too, will begin to decline. The labor force of women aged 20 to 24 is projected to have the greatest growth between 1975 and 1980, increasing by about 1.0 million. It should grow a further quarter of a million during the first half of the 1980 decade, but will then recede rapidly, dropping by nearly 700,000 by 1990 (table 1 and 2). The male labor force 20 to 24 years of age should increase by about 700,000 during the late 1970's. It is then expected to decrease by 300,000 during the early 1980's and by more than one million during the 1985-1990 period.

The completion of the entry of the "baby boom" into the labor force over the next few years will have the effect of making the overall work force younger. For example, the median age of the labor force, which already decreased from 39.6 years in 1970 to 36.0 years in 1975, is projected to drop even further to 34.9 years by 1980. Although the youth labor force will shrink after 1980, the median age of the labor force should increase only slightly during the decade, to about 36.5 years in 1990 (table 3). This is because the baby boom cohorts will still be relatively young, the oldest being only 39 in 1990. 2/

The prime age labor force

The prime age labor force--that composed of persons 25 to 54--should grow more rapidly than the other two groups over the next fifteen years. It is projected to expand at a rate of 2.1 percent a year during the seventies and at a rate of 2.4 percent per year in the early eighties, before returning to the more moderate growth rates of the seventies (table 2). The principal factor behind this projected increase is the aging of the baby boom cohorts; by 1990, all these post-World War II "babies" will be in the prime ages. A second factor is the expected continuation of the rising trend in the labor force participation of women. As a result, the proportion of the labor force accounted for by prime age workers is expected to increase from 61 percent in 1975, to 65 percent in 1985 and to about 70 percent in 1990.

As this labor force groups grows, both in absolute and relative terms, the economy should have the potential for increased productivity. Also, with a much larger proportion of the work force in ages where the probability of unemployment is lowest, the Nation's unemployment rate (ceteris paribus) should tend to be somewhat lower than in recent years. Moreover, the conventional unemployment rate should be more similar to Perry's weighted unemployment rate because of the larger proportion of the labor force composed of prime age workers. 3/

Further, the capacity of the labor force to carry social responsibilities (such as Social Security) should be enhanced during this period. One way of illustrating this is to examine the projected changes in the ratio of nonworkers to workers, generally referred to as the "economic dependency" ratio: 4/

	1975	1980	1985	1990
Dependency Ratio	125.4	114.7	111.5	111.4

As shown, the average number of nonworkers for each 100 labor force members are projected to drop from about 125 in 1975 to 111 in 1990. This decline reflects essentially the absorption of the baby boom cohorts into the labor force. (Of course, as the baby boom cohorts will begin to retire in large numbers during the late 1990's and early in the next century, the pendulum will begin to swing in the other direction. But this matter, which is of much concern in terms of the long-term health of the Social Security System, is outside the scope of this paper.)

As shown both in table 1 and 2, the projected growth of the prime age labor force between 1975 and 1990 will also be a reflection of the aging of the post World War II baby boom. The projected growth of 6.3 million in the latter seventies, will occur mainly among persons 25 to 34 years old, and the 8.1 million growth of the early eighties should be concentrated among 35 to 44 year olds. This latter age group should also exhibit the most growth in the late eighties. To the economy as a whole, the increase in the labor force of these age groups is highly favorable. However, because of the "crowding" of workers in these central age groups, the individual members of these groups are likely to encounter more competition for certain jobs, such as junior management positions, the number of which is generally related to the number of workers of all ages. 5/

A second factor contributing to the growth of the labor force among workers of prime working age is the increased participation by women, particularly those of child bearing age. Historically, participation rates for women at the primary childbearing ages have been significantly lower than both those of younger and of somewhat older women. However, this difference, particularly in relation to older women, has been getting smaller and smaller.
While the rising trend in participation among women in the most active years of childbearing is attributable partly to the fact that many have either postponed their childbearing or have decided not to have children, it is also clearly evident among those--and they are still a large number--who are mothers of very young children. 6/

Although the growth of the prime-age female labor force is not projected to be as great in the late seventies as it was during the early part of this decade, it is expected to increase again during the early eighties, when most of the baby boom cohorts will enter this age group. Altogether, the number of prime age women in the labor force is projected to rise from 21.6 million in 1975 to 32.5 million in 1990, with their proportion of the labor force increasing from 23 to 29 percent (table 3). This represents a 2.7 percent annual average rate of increase over the 1975-90 period. Despite the fact that the prime age females make up a smaller proportion of the labor force than the prime age males, they should account for roughly half of the total labor force growth during the next fifteen years (table 2).

The older labor force

In general, the number of older workers--those 55 years and over--is projected to rise somewhat in the late seventies, remain unchanged over the early eighties, and then decline during the late eighties. The participation rates of men in these age groups have been declining for some time. For example, the rate of men age 55 to 64 fell from 82 to 76 percent between 1970 and 1975, and it is projected to drop to about 70 percent by 1990. 7/ Rates for men 65 and over are also expected to continue their long run decrease. However, their labor force is still expected to increase somewhat during the late seventies because of the continued growth of their population. During the early eighties, there should be little change with decreases in the number of older men in the labor force, offset by increases in the number of older women, whose participation rates are expected to show little change. In the late eighties, the labor force 55 and over is projected to fall, and the 1990 level of 14.3 million should be only slightly higher than the 1975 average for this group. This is probably a short term phenomen which is related to a decrease in the population 55-64, that is the surviving of the babies born during the low birth years of the 1930's. These population groups are expected to increase again after 1995. 8/

Overall, the proportion of the total labor force accounted for by persons 55 and over should decline from 15 percent in 1975 to 13 percent in 1990 (table 3). The decrease in the proportion of older workers in the labor force has several implications which are difficult to quantify. On the one hand, older workers bring to the work-place job experience that cannot always be obtained in school. On the other hand, the older work force has a lower proportion of highly educated workers, thus their withdrawal will increase the educational attainment of the work force. It is therefore difficult to assess the impact of a smaller proportion of older workers on the productivity of the labor force. One probable effect will be to enhance promotion opportunities for younger members of the labor force, indeed, some of the current pressure for earlier retirement may be caused by the competition by younger workers for higher level jobs.

Footnotes

1/ These projections supersede those which were presented by Denis F. Johnston in "The U.S. labor force: projections to 1990," Monthly Labor Review, July 1975, pp. 3-13, reprinted as Special Labor Force Report 156. These labor force projections also supersede the projections presented by Ronald E. Kutscher (without demographic detail) in "Revised BLS projections to 1980 and 1985: an overview," Monthly Labor Review, March 1976, pp. 3-5. The new projections are based on the three projections of population as given in the Current Population Reports, Series, P-25, No. 601 (Bureau of the Census, 1975). Most of the discussion in this article is in terms of the Series II projection in that publication. These labor force projections incorporate the Bureau of Labor Statistics current judgments and assumptions concerning anticipated future developments in the labor force participation rates of the several age-sex groups in the population of working age (16 and over).

2/ The "baby boom' cohort" is defined as those persons born from 1950 to 1964, a period which encompasses the peak year of births, 1961. Thus, the "baby boom" cohort was 10 to 24 years of age in 1975 and will be 25 to 39 by 1990.

3/ George L. Perry, "Changing Labor Markets and Inflation," Brookings Papers on Economic Activity (3:1970), pp. 411-441.

4/ There is no standard definition of the "economic dependency" ratio. In this case, it is defined as the persons not in the labor force (i.e. including all those less than 16 years old) divided by the labor force. See Henry S. Shryock and Jacob S. Siegel and associates, The Methods and Materials of Demography, (Bureau of the Census, 1973), page 235.

5/ One of the most consistant advocates of the concept that the baby boom cohort will experience difficulties at each stage of their life cycle is Denis F. Johnston. His most recent statement of this theme is "The Aging of the Baby Boom Cohorts," Statistical Reporter, March 1976, pp. 161-165.

6/ For further discussion of recent trends in female labor force participation, see three articles on the labor force in the November 1975 Monthly Labor Review: Allyson Sherman Grossman, "Women in the Labor Force: the Early Years," pp. 3-9; Deborah Pisetzner Klein, "Women in the Labor Force: the Middle Years," pp. 10-16; Beverly Johnson McEaddy, "Women in the Labor Force: the Later Years," pp. 17-24.

7/ For discussion of topics related to declining male labor force participation rates and related topics, see Paul O. Flaim "Discouraged Workers and Changes in Employment," Monthly Labor Review, March 1973, pp. 8-16; Robert L. Stein, "Reasons for Nonparticipation in the Labor Force," Monthly Labor Review, July 1967, pp. 22-27; Joseph L. Gastwirth, "On the Decline of Male Labor Force Participation," Monthly Labor Review, November 1973, pp. 53-54 and William V. Deutermann, Jr., "Declining Labor Force Participation of Prime Age Men," (Washington, D.C.; Bureau of Labor Statistics, 1976) unpublished.

8/ Jacob S. Siegel, Demographic Aspects of Aging and the Older Population in the United States, Current Population Reports: Special Studies: Series P-23, No. 59 (U.S. Bureau of the Census, 1976).

Table 1. Civilian noninstitutional population, civilian labor force, and civilian labor force participation rate, by age and sex, actual 1970 and 1975 and projected 1980, 1985, and 1990

(Numbers in thousands)

	Civilian noninstitutional population, July 1					Civilian labor force, annual averages					Civilian labor force participation rates, annual averages (percent of population in labor force)				
	Actual		Projected			Actual		Projected			Actual		Projected		
	1970	1975	1980	1985	1990	1970	1975	1980	1985	1990	1970	1975	1980	1985	1990
BOTH SEXES															
Total, 16 years and over	137,809	151,389	163,200	171,900	178,967	82,714	92,613	101,673	108,602	113,839	60.0	61.2	62.3	63.2	63.6
16 to 24 years	30,159	34,530	36,108	33,389	30,078	17,829	22,265	24,266	23,067	20,952	59.1	64.5	67.2	69.1	69.7
25 to 54 years	70,085	75,983	82,953	91,942	101,138	50,388	56,182	62,497	70,616	78,567	71.9	73.9	75.3	76.8	77.7
55 years and over	37,563	40,876	44,139	46,569	47,751	14,497	14,165	14,910	14,919	14,320	38.6	34.7	33.8	32.0	30.0
MEN															
Total, 16 years and over	64,536	71,468	77,089	81,119	84,379	51,194	55,615	60,000	62,903	65,220	79.3	77.8	77.8	77.5	77.3
16 to 19 years	7,281	8,049	8,037	6,870	6,485	4,005	4,760	4,905	4,181	3,976	55.0	59.1	61.0	60.9	61.3
16 and 17 years	3,878	4,199	4,074	3,503	3,173	1,808	2,039	2,061	1,777	1,612	46.6	48.6	50.6	50.7	50.8
18 and 19 years	3,402	3,850	3,963	3,367	3,312	2,197	2,721	2,849	2,404	2,364	64.6	70.7	71.8	71.4	71.4
20 to 24 years	6,897	8,769	9,584	3,986	8,129	5,709	7,398	8,069	7,795	6,671	82.8	84.4	84.2	83.0	82.1
25 to 34 years	11,707	14,566	17,196	18,997	19,590	11,311	13,854	16,369	18,021	18,545	96.6	95.1	95.2	94.9	94.7
35 to 44 years	10,831	10,745	12,147	14,917	17,471	10,464	10,288	11,600	14,192	16,571	96.6	95.7	95.5	95.1	94.8
45 to 54 years	11,062	11,330	10,841	10,721	12,085	10,417	10,426	9,892	9,709	10,901	94.2	92.0	91.2	90.6	90.2
55 to 64 years	8,708	9,221	9,791	10,000	9,592	7,124	6,982	7,275	7,162	6,704	81.8	75.7	74.3	71.6	69.9
55 to 59 years	4,718	4,963	5,313	5,194	4,888	4,218	4,185	4,448	4,283	3,990	89.4	84.3	83.7	82.5	81.6
60 to 64 years	3,990	4,258	4,478	4,806	4,704	2,906	2,797	2,827	2,879	2,714	72.8	65.7	63.1	59.9	57.7
65 years and over	8,047	8,784	9,492	10,228	11,027	2,164	1,906	1,890	1,843	1,852	26.9	21.7	19.9	18.0	16.8
65 to 69 years	3,061	3,497	3,734	3,942	4,236	1,278	1,108	1,125	1,104	1,125	41.8	31.7	30.1	28.0	26.6
70 years and over	4,986	5,287	5,759	6,286	6,791	886	799	765	739	727	17.8	15.1	13.3	11.8	10.7
WOMEN															
Total, 16 years and over	73,273	79,921	86,111	90,781	94,588	31,520	36,998	41,673	45,699	48,619	43.0	46.3	48.4	50.3	51.4
16 to 19 years	7,485	8,225	8,160	7,018	6,612	3,241	4,038	4,226	3,762	3,649	43.3	49.1	51.8	53.6	55.2
16 and 17 years	3,796	4,113	3,972	3,420	3,089	1,324	1,652	1,712	1,551	1,448	34.9	40.2	43.1	45.4	46.9
18 and 19 years	3,689	4,111	4,188	3,598	3,523	1,917	2,387	2,514	2,221	2,201	52.0	58.1	60.0	61.5	62.5
20 to 24 years	8,494	9,486	10,327	10,115	8,852	4,874	6,069	7,066	7,329	6,659	57.4	64.0	68.4	72.5	75.2
25 to 34 years	12,724	15,514	18,108	19,967	20,582	5,698	8,456	10,394	12,210	13,077	44.8	54.5	57.4	61.2	63.5
35 to 44 years	11,771	11,618	13,084	15,903	18,525	5,967	6,493	7,633	9,723	11,678	50.7	55.9	58.3	61.1	63.0
45 to 54 years	11,986	12,206	11,577	11,437	12,885	6,531	6,665	6,609	6,761	7,795	54.4	54.6	57.1	59.1	60.3
55 to 64 years	9,757	10,349	11,035	11,238	10,671	4,153	4,244	4,628	4,740	4,514	42.5	41.0	41.9	42.2	42.3
55 to 59 years	5,181	5,470	5,881	5,698	5,299	2,547	2,618	2,891	2,870	2,703	49.2	47.9	49.2	50.4	51.0
60 to 64 years	4,576	4,878	5,154	5,549	5,372	1,606	1,626	1,737	1,870	1,811	35.1	33.3	33.7	33.7	33.7
65 years and over	11,051	12,521	13,820	15,103	16,461	1,056	1,033	1,117	1,174	1,250	9.6	8.2	8.1	7.8	7.6
65 to 69 years	3,814	4,433	4,748	5,028	5,419	644	640	692	721	768	16.9	14.4	14.6	14.3	14.2
70 years and over	7,236	8,088	9,072	10,075	11,042	412	392	425	453	482	5.7	4.8	4.7	4.5	4.4

Source: The 1970 and 1975 population are estimated from Current Population Reports, Series P-25, No. 614, Labor Force data for 1970 are from Special Labor Force Report 156. The 1975 labor force data are from Current Population Survey estimates. Population projections are from Current Population Reports, Series P-25, No. 601.

Table 2. Net changes in the civilian labor force 16 years old and over, by age and sex, 1970-75, 1975-80, 1980-85 and 1985-90

Sex and age groups	Net change (in thousands)				Annual average rate of change 1/ (in percent)			
	1970-75	1975-80	1980-85	1985-90	1970-75	1975-80	1980-85	1985-90
BOTH SEXES								
Total, 16 years and over	9,899	9,060	6,929	5,237	2.26	1.87	1.32	0.94
16 to 24 years	4,436	2,001	-1,199	-2,115	4.44	1.72	-1.01	-1.92
16 to 19 years	1,552	333	-1,188	-318	3.88	.74	-2.79	-.82
20 to 24 years	2,884	1,695	-38	-1,797	4.82	2.37	2/	-2.53
25 to 54 years	5,795	6,315	8,119	7,951	2.18	2.13	2.44	2.13
25 to 34 years	5,301	4,453	3,468	1,391	5.43	3.64	2.44	.90
35 to 44 years	350	2,452	4,682	4,334	0.42	2.73	4.36	3.33
45 to 54 years	143	-590	-31	2,226	0.17	-.70	2/	2.54
55 years and over	-332	745	9	-599	-0.46	1.03	2/	-.82
55 to 64 years	-51	677	-1	-684	-0.09	1.17	2/	-1.18
65 years and over	-281	68	10	85	-1.83	.46	.07	0.56
MEN								
Total, 16 years and over	4,421	4,385	2,903	2,317	1.66	1.52	0.94	0.72
16 to 24 years	2,444	816	-998	-1,329	4.49	1.30	-1.60	-2.35
16 to 19 years	755	145	-724	-205	3.45	.60	-3.19	-1.01
20 to 24 years	1,689	698	-301	-1,124	5.18	1.80	-.76	-3.11
25 to 54 years	2,376	3,293	4,061	4,095	1.42	1.82	2.04	1.86
55 years and over	-400	277	-160	-449	-.88	.61	-0.35	-1.02
55 to 64 years	-142	293	-113	-458	-.40	.82	-.31	-1.32
65 and over	-258	-16	-47	9	-2.54	-.17	-.50	.10
WOMEN								
Total, 16 years and over	5,478	4,675	4,026	2,920	3.20	2.38	1.84	1.24
16 to 24 years	1,992	1,185	-201	-786	4.39	2.22	-.36	-1.47
16 to 19 years	797	188	-464	-113	4.40	0.91	-2.33	-.61
20 to 24 years	1,195	997	263	-673	4.39	3.04	.73	-1.93
25 to 54 years	3,418	3,022	4,058	3,856	3.44	2.62	3.05	2.52
55 years and over	68	468	169	-150	.26	1.70	.58	-.51
55 to 64 years	91	384	112	-226	.43	1.73	0.48	-.98
65 years and over	-23	84	57	76	-.44	1.56	1.00	1.25

1/ Compounded continuously.

2/ Less than 0.05 percent.

Table 3. Distribution of civilian labor force, by age and sex, actual 1970 and 1975 and projected 1980, 1985, and 1990

Sex and age group	Percent distribution				
	Actual		Projected		
	1970	1975	1980	1985	1990
BOTH SEXES					
Total, 16 years and over	100.0	100.0	100.0	100.0	100.0
16 to 24 years	21.6	24.0	23.9	21.2	18.4
16 to 19 years	8.8	9.5	9.0	7.3	6.7
20 to 24 years	12.8	14.5	14.9	13.9	11.7
25 to 54 years	60.9	60.7	61.5	65.0	69.0
25 to 34 years	20.6	24.1	26.3	27.8	27.8
35 to 44 years	19.9	18.1	18.9	22.0	24.8
45 to 54 years	20.5	18.5	16.2	15.2	16.4
55 years and over	17.5	15.3	14.7	13.7	12.6
55 to 64 years	13.6	12.1	11.7	11.0	9.9
65 years and over	3.9	3.2	3.0	2.8	2.7
Median age	39.6	36.0	34.9	35.3	36.5
MEN					
Total, 16 years and over	61.9	60.1	59.0	57.9	57.3
16 to 24 years	11.7	13.1	12.8	11.0	9.4
16 to 19 years	4.8	5.1	4.8	3.8	3.5
20 to 24 years	6.9	8.0	8.0	7.2	5.9
25 to 54 years	38.9	37.3	37.2	38.6	40.4
55 years and over	11.2	9.6	9.0	8.3	7.5
55 to 64 years	8.6	7.5	7.2	6.6	5.9
65 years and over	2.6	2.1	1.9	1.7	1.6
Median age	37.6	36.7	35.5	35.9	36.9
WOMEN					
Total, 16 years and over	38.1	39.9	41.0	42.1	42.7
16 to 24 years	9.8	10.9	11.1	10.2	9.1
16 to 19 years	3.9	4.4	4.2	3.5	3.2
20 to 24 years	5.9	6.6	6.9	6.7	5.8
25 to 54 years	22.0	23.3	24.2	26.4	28.6
55 years and over	6.3	5.7	5.7	5.4	5.1
55 to 64 years	5.0	4.6	4.6	4.4	4.0
65 years and over	1.3	1.1	1.1	1.1	1.1
Median age	39.5	34.9	34.1	34.6	35.7

APPLICATION OF AN ITERATIVE REGRESSION TECHNIQUE TO A NATIONAL POTENTIAL CROPLAND SURVEY

J. Jeffery Goebel, Iowa State University

The problem of estimation in a sample survey when data are available from outside sources for several characteristics of the population is considered. The control totals are incorporated into the estimation through an iterative generalized least squares regression technique that is applicable for any sampling scheme. We apply the technique to a national survey concerned with estimating the acreage of potential cropland and the difficulties involved in developing the land into cropland. This study was motivated by a desire to establish our future agricultural capabilities and to identify programs that might be required to ensure a supply of food sufficient to meet national and international demands. The multi-stage sampling scheme employed systematic and cluster sampling. The goal of the design was to minimize the variances of the national estimates, subject to a fixed cost restriction and certain accuracy restrictions on regional estimates.

* * * * *

We consider the case where n multivariate and categorical (or discrete) observations are taken--the n observed units are selected by a multi-stage procedure that does not necessarily give each unit an equal probability of being included in the sample. We know the sampling procedure (including the probabilities of selection) and also several of the population figures for some of the categories. In other words, we assume our observations can be represented by an r-dimensional classification, where each dimension represents an attribute and the i-th attribute (or classification) has s_i categories. Let $P_{j_1, j_2, \ldots, j_r}$ be the proportion of the population whose first classification is j_1, whose second classification is $j_2, \ldots$, and whose r-th classification is j_r, where $j_i = 1,2,\ldots,s_i$ and $i = 1, 2,\ldots,r$. Then we know some of the subtotals such as the fraction of the population in category j_1.

It certainly seems desirable to include such information in the inference procedure. This has been recognized for a long time [see, for example, El-Badry and Stephan (1955)], and becomes increasingly relevant as the stockpile of data swiftly grows. Using the extra data should improve the estimates one wishes to derive from the sample. Not only can one significantly reduce the variances (as we have found in a separate study), but also the figures to be presented will match with other published figures.

Our situation is related to contingency table estimation with known column and row totals. Deming and Stephan (1940) were one of the first to consider this situation, where the $P_{i.}$'s and $P_{.j}$'s are specified and estimates for the cell entries are desired.

	1	2	. . .	s	Total
1	$\tilde{P}_{11}$	$\tilde{P}_{12}$	. . .	$\tilde{P}_{1s}$	$P_{1.}$
2	$\tilde{P}_{21}$	$\tilde{P}_{22}$	. . .	$\tilde{P}_{2s}$	$P_{2.}$
⋮	⋮	⋮		⋮	⋮
r	$\tilde{P}_{r1}$	$\tilde{P}_{r2}$	. . .	$\tilde{P}_{rs}$	$P_{r.}$
Total	$P_{.1}$	$P_{.2}$	. . .	$P_{.s}$	$P_{..}$

They assumed the $\{n_{ij}\}$ were multinomially distributed and minimized $S = \sum_{i=1}^{r} \sum_{j=1}^{s} w_{ij}(n_{ij} - n\tilde{P}_{ij})^2$ over $\{\tilde{P}_{ij}\}$ subject to the marginal restrictions $\sum_{i=1}^{r} \tilde{P}_{ij} = P_{.j}$ and $P_{i.} = \sum_{j=1}^{s} \tilde{P}_{ij}$, using $w_{ij} = \frac{1}{n_{ij}}$.
They obtained the closed form solution

$$\tilde{P}_{ij} = \hat{p}_{ij}(1 + \lambda_{i.} + \lambda_{.j}) ,$$

where $\hat{p}_{ij} = n_{ij}/n$ and the λ's are obtained from simultaneous equations in the n_{ij}'s and controls; they also gave a quick simple iterative technique for construction of $\{\tilde{P}_{ij}\}$. Stephan (1942) improved on the iterative procedure. J. H. Smith (1947), under the multinomial assumption, developed a maximum likelihood estimator: $\hat{P}_{ij} = \hat{p}_{ij} (a_i + b_j)^{-1}$, where $\{a_i\}$ and $\{b_j\}$ are functions of the controls, $\{P_{i.}\}$ and $\{P_{.j}\}$.

Others have used different criteria for constructing estimators. El-Badry and Stephan (1955) derived generalized least squares estimates which are approximately equivalent to the maximum likelihood estimators. Ireland and Kullback (1968) proposed estimating the cell probabilities of the contingency table by using the theory of minimum discrimination information, [i.e., the discrimination information $I(\cdot) = \Sigma\Sigma P_{ij} \ln (P_{ij} / \hat{p}_{ij})$ is minimized]. Their estimates are iterative (and shown to converge) and best asymptotically normal. They also discussed higher dimensional (>2) contingency tables with various marginals known.

More recently Chen and Fienberg (1974) obtained iterative MLEs for the case where only one of the classifications is known for some of the observations (and controls are present). One can inject other types of constraints--either instead of the types of controls to be discussed here or in conjunction with them. Bishop, Fienberg, and Holland (1975) in their new book Discrete Multivariate Analysis discuss these additional constraints, plus many other facets of contingency table analysis.

Grizzle, Starmer, and Koch (1969) analyzed categorical data by linear models--hence having all the advantages of using weighted regression (the estimates, testing, etc.). They considered taking $\{n_{i.} : i = 1,2,\ldots,s\}$ samples from s multinomial distributions, each having r categories of response, and gave a noniterative procedure.

Let us now consider regression approaches in the estimation of categorical data. Suppose we have weights, say $\{a_i\}$, for each observation such that the weighted sample mean

$$\bar{y}_w = \frac{1}{n} \sum_{i=1}^{n} a_i y_i$$

is unbiased for $\bar{Y}$ = the population mean of the characteristic of interest, Y, where the $\{a_i\}$ are nonnegative and add to n . If k auxiliary variables are available (they can be either continuous, or "0-1," as in our case) and the population figures (means or totals) are known for them, we can then use the generalized regression estimator

$$\bar{y}_G = \bar{y}_w + (\bar{\underset{\sim}{X}} - \bar{\underset{\sim}{x}}_w)\underset{\sim}{b}_G , \quad (1)$$

where

$$\underset{\sim}{b}_G = [\Sigma\, a_i\, \underset{\sim}{Z}_i \underset{\sim}{Z}_i']^{-1} [\Sigma\, a_i y_i\, \underset{\sim}{Z}_i] ,$$

$$\underset{\sim}{Z}_i' = (1, X_{i1} - \bar{x}_{w1}, X_{i2} - \bar{x}_{w2}, \ldots, X_{ik} - \bar{x}_{wk}),$$

and $\bar{\underset{\sim}{x}}_w$ is the vector of the weighted sample means of the k+1 X's, with all $X_{i0} = 1$. Equation (1) can be written as a linear function of the observed y_i's, say

$$\bar{y}_G = \sum_{i=1}^{n} w_i y_i . \quad (2)$$

Note that under simple random sampling all the a_i are 1 and $\bar{y}_G$ is simply the ordinary regression estimator.

The actual algorithm we use is slightly different than this form--it includes differences from population means (instead of from the weighted sample means) and an additional initial "weight." The first-stage regression weight is $w_i^{(0)}$, where for i = 1,2,...,n,

$$w_i^{(j)} = a_i(\frac{1}{n} + u_i^{(j)})(1 + n\,\bar{u}_w^{(j)})^{-1} , \quad (3)$$

$$u_i^{(j)} = g_i^{(j)}\, (\underset{\sim}{X}_i - \bar{\underset{\sim}{X}})'\, [\underset{\sim}{A}_j]^{\dagger}\, (\bar{\underset{\sim}{X}} - \bar{\underset{\sim}{x}}_w),$$

$$\underset{\sim}{A}_j = \sum_{t=1}^{n} a_t\, g_t^{(j)}\, (\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})',$$

$[\underset{\sim}{A}_j]^{\dagger}$ denotes the generalized inverse of $\underset{\sim}{A}_j$, and $\underset{\sim}{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ik})'$ is now k-dimensional. The $\{g_t^{(0)}\}$ are the additional set of weights mentioned above and are commonly all unity. There are some situations where improvements can be made by using nonconstant $g_t^{(0)}$'s. An example is stratified sampling, where the means are known only for the population, and not by strata. The weighted combined estimator given by Cochran (1963) is such that a_t is inversely proportional to the sampling fraction f_h for the stratum and

$$g_t^{(0)} = N_h(1 - f_h)/(n_h - 1),$$

where the t-th observation is from stratum h .

If $|u_i^{(j)}| > M/n$ for any i (where $.1 \leq M \leq 1$ is fixed), we iterate--do a $(j+1)^{st}$ step. For each distance $d_i^{(j)} = |4n\, u_i^{(j-1)}/(3M)|$ an adjusting weight $g_{(j)i}$ is computed as:

$$g_{(j)i} = \begin{cases} 1 & , 0 \leq d_i^{(j)} < \frac{1}{2} \\ 1 - \frac{4}{5}(d_i^{(j)} - \frac{1}{2})^2 & , \frac{1}{2} \leq d_i^{(j)} \leq 1 \\ 4/(5d_i^{(j)})^{-1} & , 1 < d_i^{(j)} . \end{cases}$$

Then we set $g_i^{(j)} = \prod_{h=0}^{j} g_{(j)i}$ and use line (3) to compute a new set of $w_i^{(j)}$'s .

Iteration will cease when:

(i) the $\{u_i^{(j)}\}$ are such that $|u_i^{(j)}| > M/n$ for all "i," since then the distances $d_i^{(j)}$ are too large and the required limitations on the weights cannot be met; or

(ii) if a specified number of iterations have occurred; or

(iii) if $|u_i^{(j)}| < M/n$ for all "i."

The third condition is the desired reason for ceasing iteration, and the result will be regression weights with the following properties:

$$w_i \geq 0 , \; i = 1,2,\ldots,n;$$

$$(1\text{-}M) \max_i\{w_i/a_i\} \leq (1+M) \min_i\{w_i/a_i\};$$

$$\sum_{i=1}^{n} w_i\, \underset{\sim}{X}_i = \bar{\underset{\sim}{X}} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k)';$$

$$\sum_{i=1}^{n} w_i = 1 .$$

Also, the associated regression coefficients will be best asymptotically normal estimates.

Let us now consider a recently completed national survey which was analyzed using this regression technique. In early 1975 there was considerable national and international concern about the global food situation. An example is the following excerpt from an editorial in the Journal of Soil and Water Conservation: "Crucial to our nation's ability to provide increased food supplies is the availability of our agricultural resource base. ...Now as we look to the future, particularly in terms of world food trade needs and world food security, we must be concerned about the adequacy of our land resource base." The United States Department of Agriculture was receiving numerous requests for data on potential cropland from Congress and from others. Data on current land use and on the potential for new cropland were needed so that the Department of Agriculture could respond accurately to these inquiries. In spring 1975 the Statistical Laboratory at Iowa State University cooperated with the Soil Conservation Service and the Economic Research Service of the United States Department of Agriculture in designing a national sample for this purpose.

At the time of design, the types of data desired for the sample segments of land were:

(i) type of soil;
(ii) 1975 land use--cropland, pasture and range, forest, other land, urban, and water;
(iii) 1967 land use;
(iv) types of development problems that would significantly inhibit development for cropland;
(v) type of development necessary for conversion to cropland;
(vi) the potential for conversion to cropland within the next 10-15 years.

Later a seventh variate was added:

(vii) whether or not the sample point is prime farmland.

The seven variables are all categorical. Items (i) and (iii) had been previously estimated. Acreages classified by "land capability unit class and subclass" and by "1967 land use" were available for each state from the 1967 Conservation Needs Inventory [see reference (12)]. Therefore, it was decided to use these data as "control" data. Items (ii) and (iv) through (vii) were collected on the sample sites.

The problem of sampling the nation to obtain acreage estimates has been studied--and a sample existed. However, the existing sample was designed to provide answers at the county level. Time and money were not available to study potential cropland in such detail and a new sample was designed. Usable estimates were desired for all 50 states plus Puerto Rico and the Virgin Islands. Costs dictated a national sample of about 500 counties. Each selected county was given roughly the same number of sample points. This was to equalize the work load because regular SCS field personnel were to do the work. Also, it was necessary to specify the number of sample counties per state immediately so that funds could be dispersed. The allocation per state was based upon

(i) the number of counties in the state,
(ii) the size of the state, and
(iii) the acreage in cropland in 1967.

For example, Massachusetts was given five counties, Illinois and Iowa 16 each, while 28 counties were selected in Texas.

A brief synopsis of the sampling scheme follows. The universe for this study consisted of the 1.44 billion acres (in the 50 states, Puerto Rico, and the Virgin Islands) considered to be in inventory for the 1967 Conservation Needs Inventory [see reference (12)]. Inventory acres were basically all rural land, except for federal land not cropped. Within each state (plus Puerto Rico and the Virgin Islands), counties served as clusters and were selected on a systematic basis. Some of the 506 counties included in the Potential Cropland Survey were divided into substrata due to their heterogeneity. Secondary units were then selected within substrata within counties using a systematic scheme. A total of 5,300 secondary units were selected. The secondary units are square areas of land, typically 160 acres in size. The major exceptions are: (i) in the northeastern states, 100 acre units were used; and (ii) in the western and mountain regions, some 40 acre squares were used in irrigated areas, while 640 acre units were used in some nonirrigated areas. The ultimate sampling units were points selected on a two-dimensional systematic basis within secondary units. As discussed, for example, by Strand and Huang (1973), this use of points does not seem to appreciably affect the accuracy of acreage estimates relative to completely observing and mapping the 160 acre secondary units.

For the survey on potential cropland, the regression weights technique was applied to data for each state. Some combining of sample points was done in an effort to reduce computational costs. In effect "new observations" were formed that combined the information by soil grouping (land capability class and subclass), 1967 land use, and substrata within counties. Note that acreages for the soil groupings and 1967 land uses serve as the control figures, while the substrata within counties are geographical regions where homogeniety is expected. The $\{a_t\}$ were constant for each substrata within a county and depended upon the sampling rates.

Two tables of national estimates are included as examples of estimates produced. Of particular interest are: (i) the 7% decrease from 1967 to 1975 in land classified as cropland; and (ii) that cropland acreage could be increased by at least a fourth if all land with high and medium potential was converted to cropland. Variance estimates are being computed for the state, regional, and national estimates derived for this survey.

ACKNOWLEDGMENT

This research was supported in part through Cooperative Agreement 12-10-001-92 between Iowa State University and the Soil Conservation Service, United States Department of Agriculture.

REFERENCES

(1) Bishop, Yvonne M. M.; Fienberg, Stephen E.; and Holland, Paul W. (1975). Discrete Multivariate Analysis. The MIT Press: Cambridge, Mass.
(2) Chen, Tar and Fienberg, Stephen E. (1974). "Two-dimensional contingency tables with both completely and partially cross-classified data." Biometrics 30, 629-641.
(3) Cochran, William G. (1963). Sampling Techniques. John Wiley and Sons: New York.
(4) Deming, W. E. and Stephan, F. F. (1940). "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known." Ann. Math. Statist. 11, 427-444.
(5) El-Badry, M. A. and Stephan, F. F. (1955). "On adjusting sample tabulations to census counts." J. Amer. Statist. Assoc. 50, 738-762.
(6) Grizzle, J. E.; Starmer, C. F.; and Koch, G. G. (1969). "Analysis of categorical data by linear models." Biometrics 25, 489-504.
(7) Huang, Elizabeth; Goebel, J. Jeffery; and Fuller, Wayne A. (1974). "Regression M-weights: Computer algorithm." Statistical Laboratory Survey Section, Iowa State University: Ames, Iowa.
(8) Ireland, C. T. and Kullback, S. (1968). "Contingency tables with given marginals." Biometrika 55, 179-188.
(9) Smith, J. H. (1947). "Estimation of linear functions of cell proportions." Ann. Math. Statist. 18, 231-254.
(10) Stephan, F. F. (1942). "An iterative method of adjusting sample frequency tables when expected marginal totals are known." Ann. Math. Statist. 13, 166-178.
(11) Strand, Norman V. and Huang, Her Tzai (1973). "Conservation Needs Inventory: Influence of sample size and form of estimator on sampling variation in selected counties in Iowa, 1957 and 1967." Statistical Laboratory Survey Section, Iowa State University, and Soil Conservation Service: Ames, Iowa.
(12) U. S. Department of Agriculture (1971). Basic Statistics--National Inventory of Soil and Water Conservation Needs, 1967. Statistical Bulletin 461.

TABLE 1: CHANGES IN LAND USE: 1967 VERSUS 1975 (ESTIMATED ACRES x 1,000)

Land use in 1967	Land use in 1975						
	Cropland	Pasture and range	Forest	Other land	Urban	Water	Total
Cropland	351,651	52,884	8,265	12,977	4,846	618	431,241
Pasture and range	31,907	442,352	14,096	14,178	3,211	1,111	506,855
Forest	11,027	62,469	348,681	15,801	4,423	2,152	444,553
Other land	5,832	13,176	4,406	26,874	4,156	2,827	57,271
Total	400,417	570,881	375,448	69,830	16,636	6,708	1,439,920

TABLE 2: 1975 STATUS OF 1967 CNI ACREAGE BY LAND CAPABILITY CLASS AND SUBCLASS (ESTIMATED ACRES x 1,000)

Soil type (class & subclass)	Potential for cropland of 1975 pasture, range, forest, and other land				Cropland	Urban and water	Total
	High	Medium	Low	Zero			
1-	5,091	343	3,002	1,889	33,389	1,301	45,015
2E	19,987	3,750	25,486	8,194	87,594	3,151	148,162
2W	9,545	2,069	13,420	6,243	60,116	1,789	93,182
2S	2,117	268	2,652	1,211	20,451	625	27,324
2C	1,914	730	7,378	920	19,709	55	30,706
3E	17,239	8,605	56,455	16,564	70,351	2,035	171,249
3W	6,022	2,574	26,963	10,194	31,377	2,913	80,043
3S	2,357	1,257	9,034	2,706	11,455	320	27,129
3C	1,215	268	1,787	112	9,662	50	13,094
4E	5,866	4,385	52,243	20,564	29,693	1,795	114,646
4W	1,675	1,569	16,282	11,193	3,968	776	35,463
4S	842	767	12,719	7,181	5,904	728	28,141
4C	0	338	1,062	149	334	0	1,883
5W	476	1,992	14,319	9,779	1,497	892	28,955
5S	5	0	1,927	84	32	28	2,076
6E	3,202	2,394	33,313	123,635	8,794	1,511	172,849
6W	40	153	3,610	7,227	562	103	11,695
6S	442	1,084	21,087	51,850	3,502	444	78,409
6C	231	259	7,793	2,504	362	71	11,220
7-8	0	0	0	312,257	1,665	4,757	318,679
Total	78,266	32,805	310,532	594,556	400,417	23,344	1,439,920

STRATIFICATION INDEX: METHODOLOGY & ANALYSIS (CANADIAN LFS)

G.B. Gray, Statistics Canada

Introduction

Large scale surveys, such as the Canadian LFS, the Current Population Survey and many others utilize stratification in setting up the sample frame. This means that, instead of drawing a sample only to represent the whole population, separate independent samples are drawn to represent sub-populations called strata, which may be counties, states, provinces, or cities in the case of area samples (of which the Canadian LFS is a typical example), or lists of establishments in a certain industry group, size groups according to the number of employees or assets in the case of list samples.

To paraphrase Kish [6], there are 3 basic reasons for stratifications and these include:

i) to decrease the variances of the sample estimates,
ii) to employ different methods and procedures within them, and
iii) to employ the sub-populations representing the strata as separate domains of study.

The first reason leads to what may be called optimal stratification whereby the maximum proportion of the MSE between strata is removed so as to minimize the sampling variances within strata. The second and third lead to what one might call administrative stratification whereby special administrative procedures must be undertaken in certain sub-populations.

In this paper, we will concentrate on the methodology in which the sampling variance of characteristic totals (or means) is derived when strata are delineated compared with the sampling variances when they are not delineated.

The Canadian LFS is a typical area sample which has been stratified at several levels: (a) province, (b) type of area within province (NSRU and SRU), (c) economic regions, (d) groups of urban and rural enumeration areas within economic regions, (e) census metropolitan areas and large cities and (f) subunits within met areas and large cities.

Case (d) is the only one that belongs to the optimal stratification category. Case (f) could have been undertaken by optimal stratification but in practice was not because of the time factor and complexities involved. Furthermore, growth and characteristic changes in met areas tend to negate the advantages of optimal stratification while rural areas and small towns tend to remain stable over time where time in the future is of primary importance in continuous surveys.

Whatever the reason for stratification, it is desirable to determine whether or not we are getting our money's worth in effecting a significant reduction in the sampling variance as a result of stratification. A stratification index, measuring the fractional reduction in the variance because of stratification is developed and some empirical examples based on the Canadian LFS (10 months' data from March-December 1975 just prior to redesign) are presented.

Development of Stratification Index

(a) Simple Estimate

$$\hat{X} = \sum_{h=1}^{L} \hat{X}_h, = \sum_{h=1}^{L} \sum_i{}' \frac{1}{n_h p_{i/h}} \hat{X}_{hi}, \text{ where}$$

$\hat{X}_{hi}$ = estimate of characteristic total in psu i, stratum h

n_h = no. of selected psu's out of N_h in stratum h

$p_{i/h}$ = relative size of psu i in stratum h

(b) Variance of Simple Estimate (with stratification)

$$V(\hat{X}) = \sum_{h=1}^{L} V(\hat{X}_h), \text{ where } V(\hat{X}_h) \text{ may be stated in several ways, two of which are:}$$

i) $$V(\hat{X}_h) = \sum_{i<i'} (\pi_{i/h}\,\pi_{i'/h} - \pi_{ii'/h})(X_{hi}/\pi_{i/h} - X_{hi'}/\pi_{i'/h})^2 + \sum_{i=1}^{N_h} \sigma^2_{hi}/\pi_{i/h}, \text{ where}$$

Yates-Grundy [8]

$\pi_{i/h}$ and $\pi_{ii'/h}$ respectively denote individual inclusion probability of unit h and joint inclusion probability of i and i'. Note that $\pi_{i/h} = n_h p_{i/h}$ and

ii) $$V(\hat{X}_h) = N_h^2\, \sigma_h^2/n_h [1 + (n_h - 1) r_{FP:h}] + \sum_{i=1}^{N_h} \sigma^2_{hi}/n_h p_{i/h}, \text{ where}$$

σ_h^2 = pop'n variance between psu's

$r_{FP:h}$ = finite pop'n correlation.

Here, $$\sigma_h^2 = \sum_i p_{i/h} \left(\frac{X_{hi}}{N_h p_{i/h}} - \frac{X_h}{N_h}\right)^2$$ Gray [3] & Sukhatme [7]

and

$$r_{FP:h}\, \sigma_h^2 = \frac{1}{n_h(n_h-1)} \sum_{i\neq i'} \pi_{ii'/h} \left(\frac{X_{hi}}{N_h p_{i/h}} - \frac{X_h}{N_h}\right) \left(\frac{X_{hi'}}{N_h p_{i'/h}} - \frac{X_h}{N_h}\right)$$

if sampling without replacement has occurred, and $r_{FP:h} = 0$ if sampling with replacement has occurred.

(c) Ratio Estimate

$$\hat{X} = \sum_{a=1}^{A} P_a (\hat{X}_a/\hat{P}_a) = \sum_{a=1}^{A} P_a \hat{R}_a, \text{ where}$$

P_a = independent source data of category a (eg. age-sex pop'n), and

$\hat{X}_a/\hat{P}_a = \hat{R}_a$ = est'd ratio of characteristic among category a. Then in variance and covariance formulas, X_{hi} may be replaced by

$\sum_a (X_{hia} - R_a P_{hia})$, where $R_a = E\hat{X}_a/E\hat{P}_a$

$= X_a/P_a$

Suppose now that instead of an area delineated into L strata, we have no stratification in the area but simply select $n = \Sigma_{nh}$ psu's with or without replacement with pps and sub-sample within in the same manner.

(d) Then,

$$V_{\bar{S}}(\hat{X}) = N^2\sigma^2/n\cdot[1 + (n-1)r_{FP}] + \sum_{i=1}^{N} \sigma^2_{hi}/np_h p_{i/h}, \text{ where}$$

the same psu delineation is assumed to have occurred and where $n = \sum_h n_h$ psu's have been selected from $N = \sum_h N_h$ psu's with or without replacement.

r_{FP} is the finite population correlation for the sampling without replacement scheme that would have been undertaken over the h strata (eg., systematic pps with units in random order).

If srs is applied $r_{FP} = -1/(N-1)$

If sampling with replacement, $r_{FP} = 0$.

$$\sigma^2 = \sum_h \sum_i p_h p_{i/h} \left(\frac{X_{hi}}{N\, p_h p_{i/h}} - \frac{X}{N}\right)^2 \quad \text{and}$$

$$= \sum_h \sum_i (X_{hi} - X/N)^2 \text{ if srs is applied.}$$

(e) Difference Between $V(\hat{X})$ and $V_{\bar{S}}(\hat{X})$

In order to derive an expression for $V_{\bar{S}}(\hat{X}) - V(\hat{X})$, the relationship between σ^2 and σ_h^2's must be derived.

By an adaptation of Sukhatme [7], it can readily be shown that

$$N^2\sigma^2 = \sum_{h=1}^{L} \frac{1}{p_h} N_h^2\sigma_h^2 + L^2\sigma^2_{BS}, \text{ where}$$

σ^2_{BS} = between stratum MSE,

$$\text{and } L^2\sigma^2_{BS} = \sum_{h=1}^{L} p_h \left(\frac{X_h}{p_h} - X\right)^2.$$

By substituting the above relationship, we find that: $V_{\bar{S}}(\hat{X}) - V(\hat{X})$ can be partitioned into 3 distinct components;as follows:

i) $\frac{L^2\sigma^2_{BS}}{n}[1 + (n-1)r_{FP}]$ ... effect of M.S.E. between strata,

ii) $\sum_{h=1}^{L} \left(\frac{n_h}{np_h} - 1\right)V(\hat{X}_h)$ ... effect due to different size strata and/ or different no. of selected psu's per stratum,

and

iii) $\sum_{h=1}^{L} \frac{N_h^2\sigma_h^2}{n_h}[(n-1)r_{FP} - (n_h-1)r_{FP:h}]$

... effect of different f.p.c.'s with stratification (would be zero if sampling undertaken with replacement).

The stratification index is defined by $[V_{\bar{S}}(\hat{X}) - V(\hat{X})]/V_{\bar{S}}$ and the main component in the difference is $L^2\sigma^2_{BS}/n$ and we shall assume that $V_{\bar{S}}(\hat{X}) - V(\hat{X}) \doteq L^2\sigma^2_{BS}/n$.

(f) Estimates of Variance and Stratification Index

A couple major square deviation expressions are available for estimation purposes

$$\sum_{h=1}^{L} \frac{n_h}{n_h-1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{1}{n_h}\hat{X}_h\right)^2 \quad \text{estimates}$$

$$\sum_{h=1}^{L} [N_h^2\sigma_h^2/n_h \cdot(1-r_{FP:h}) + \sum_{i=1}^{N_h} \sigma^2_{hi}/n_h p_{i/h}]$$

instead of the true variance $V(\hat{X})$; however, since $r_{FP:h}$ is usually negative in most ppswor schemes, the expression usually gives a slight over-estimate of the variance.

Another expression which brings in the MSE between strata, vis.,

$$\sum_{h=1}^{L} p_h \left(\frac{\hat{X}_h}{p_h} - \hat{X}\right)^2 \text{ estimates } L^2\sigma^2_{BS} +$$

$$\sum_{h=1}^{L} \left(\frac{1}{p_h} - 1\right)V(\hat{X}_h)$$

so that a slightly biased estimate of $L^2\sigma^2_{BS}$ may be obtained.

The stratification index may be estimated in two stages, as follows:

$$I' = \frac{1}{n} \frac{\sum p_h \left(\frac{\hat{X}_h}{p_h} - \hat{X}\right)^2 - \sum_{h=1}^{L}\left(\frac{1}{p_h} - 1\right)\hat{V}(\hat{X}_h)}{\hat{V}(\hat{X})}$$

and the stratification index I is finally estimated by:

$\hat{I} = I'\hat{V}(\hat{X})/[\hat{V}(\hat{X}) + I'\hat{V}(\hat{X})] = I'/(1 + I')$, an index which can never exceed one although the estimate $\hat{I}$ could be negative.

A composite index $\hat{\bar{I}}$ over several sub-population domains may be obtained by summing the numerators and denominators of the individual $\hat{I}$'s.

TABLE I: Stratification Indexes Estimated by Levels of Stratification and Sub-populations (Canadian LFS data) and Guide to Tables II to IV

8 Characteristics studied:

Employed (Emp); Unemployed (Unemp); Employed: Agriculture (Emp Ag); Employed Non-agriculture (Non-ag); Employed: Manufacturing (Manuf); Employed: Construction (Constr); Employed: Transportation and Public Utilities (TPU) and Employed: Trade (Trade).

For the above characteristics, the following individual and composite indexes of stratification were calculated for each of 10 months (March - December, 1975) and averaged.

TABLE	SUB-POPULATION	INDEX AND STRATIFICATION	COMPOSITE INDEX
II	Province p, type of area T,	$\hat{I}_{pT1}$, by Economic Region E and deeper strata h's	$\hat{\hat{I}}_{T1}$ (Canada by type of area; over provinces p)
III	Province p, NSRU areas and econ. region E	$\hat{I}_{p2E}$, by strata h's (not included in Table II)	$\hat{\hat{I}}_{p2}$ (province NSRU, over econ. regions E)
IV	Met Area M	$\hat{I}_{M3}$ by sub-units h's within met area M	$\hat{\hat{I}}_{3}$ (for area covered by 10 met areas as indicated in table)

Type of area (self- and non-self representing unit areas or SRU/NSRU)

Observations and Conclusions

The economic regions were used as primary strata across Canada in the NSRU areas while the Census metropolitan areas and large cities were used as primary strata in SRU areas. If an economic region's population was small enough, the NSRU portion of an ER contained one stratum. Otherwise, it was sub-divided into 2 to 5 smaller strata using 3 major employed by industry groups as the stratifying variables. The met areas and large cities were divided into so-called sub-units which are also strata. These were delineated on maps containing blocks and block faces with dwelling counts. Contiguous sub-units were thus formed by drawing boundaries in areas of approximately equal dwelling counts noting the census tract boundaries and the potential for growth. Since optimal stratification was not applied in SRU areas, one would expect smaller reductions in the sampling variance as a result of stratification in SRU areas than as a result of stratification in NSRU areas.

In Table II, stratification first be economic regions within province NSRU areas and second, deeper strata within reduced the variance from that which would occur without stratification within province - NSRU areas between 10% and 44% for 7 of the 8 characteristics, the greatest reductions exceeding 40% for Employed: Agriculture and Employed: Manufacturing. In province SRU areas primary stratification by met areas or large cities and secondary stratification by sub-units within reduced the variance 4 to 17% for the same 7 characteristics from the variance with no stratification. In the same 7 of 8 cases, the exception being Employed: Trade, the summary stratification index for Canada NSRU areas was higher than for Canada SRU areas. At province type of area levels, the index was higher in the NSRU areas than in the SRU areas in 5 to 9 out of 10 provinces, depending upon the characteristic.

In Table III, where each primary stratum was the NSRU portion of each economic region rather than of each province, as in Table II,the Canada summary index was lower than the corresponding index of Table II for 5 out of 8 characteristics and the reduction varied between 2.5% and 26.5% vs. 9.3% and 43.7%, excluding Employed: Trade. The reductions are expected to be lower since Table II results include the effect of stratifying by economic regions as well as within economic regions while Table III includes the effect of only stratifying within economic regions.

Finally, in Table IV, the effect of delineating sub-units within 10 metropolitan areas comprising about 2/3 of the whole SRU area of Canada was considered. The summary indexes over the 10 met areas were not too different from those of Canada SRU in Table II. The comparisons between Tables II and IV are somewhat blurred because Table II considers the whole Canada SRU area while Table IV only the 10 major met areas. Most of the high indexes occurred among Employed: Agriculture which may be concentrated in certain fringe area sub-units. Finally, in Ottawa and Quebec City a high stratification index indicates that Employed: Manufacturing is concentrated in certain districts and hence sub-units of these two cities. Apart from these cases, the variance reductions did not appear substantial.

Conclusion

Stratification resulted in significant reductions in the sampling variance in Employed: Agriculture and Employed: Manufacturing - perhaps 15% to 20% at the primary stratification stage of delineating economic regions (obtained by subtracting $\bar{I}_{T2}$ from $\bar{I}_{T1}$) and another 20% to 25% through deeper stratification within economic regions. In the case of Employed, the comparable results are about 10% and 18% while for Unemployed they are only about 6% and 6%. These percentages are rough estimates because of the relatively few degrees of freedom available to estimate the MSE between strata and perhaps because of non-normal deviations. In the SRU areas the reductions are not nearly so striking.

Even if stratification results in small reductions in the variance, it does not mean that it should not be employed. As pointed out in the introduction, there are other reasons for delineating the universe into strata, not the least of which is to monitor a sample control in compact areas and perhaps redesign the sample in small areas without affecting the universe as a whole.

References and Bibliography

[1] Fellegi, I.P., "Sampling With Varying Probabilities Without Replacement: Rotating and Non-Rotating Samples", Journal of the American Statistical Association; Vol. 58 (1963), pp. 183-201: a description of method used to select psu's in NSRU areas (prior to 1971 redesign).

[2] Fellegi, I.P., Gray, G.B., and Platek, R., "The New Design of the Canadian Labour Force Survey", Journal of the American Statistical Association, Vol. 62 (1967), pp. 421-453: a description of the sample design based on 1961 Census.

[3] Gray, G.B., "Components of Variance Model in Multi-Stage Stratified Samples", Household Surveys Development Staff Survey Methodology, Vol. 1, No. 1 (June 1975), pp. 27-43.

[4] Gray, G.B., "Stratification Index: Methodology and Analysis", submitted to Household Surveys Development Staff Survey Methodology journal for possible publication.

[5] Hansen, M.H., Hurwitz, W.N., and Madow,W.G., Madow, W.G., "Sample Survey Methods and Theory", Vols. 1 and 2, New York: John Wiley & Sons (1953).

[6] Kish, L., "Survey Sampling", New York: John Wiley & Sons (1965), pp. 75-77.

[7] Sukhatme, P.V., "Sampling Theory of Surveys With Applications", The Iowa State University Press, Ames Iowa (1960), pp. 132-137.

[8] Yates, F., and Grundy, P.M., "Selection Without Replacement From Within Strata With Probability Proportional To Size", Journal Of The Royal Statistical Society, Series B, Vol. 15 (1953), pp. 253-261.

TABLE II: $\hat{I}_{pT1}$ by Province and Type of Area $\hat{\bar{I}}_{T1}$ for Canada and Type of Area, by Characteristic (Mar-Dec, 1975 Averages)

	Characteristic							
Sub-pop	Emp	Unemp	Emp Ag	Non-Ag	Manuf	Constr	TPU	Trade
Nfld SRU	.112	.110	.022	.101	.205	.079	.263	.090
Nfld NSRU	.212	.177	.016	.180	.244	.351	-.068	-.081
PEI SRU	-.013	-.006	.079	.007	.054	.032	-.031	-.010
PEI NSRU	.250	.392	-.152	.000	.327	.071	.061	-.229
NS SRU	.146	.030	.056	.196	.172	.032	.123	.085
NS NSRU	.354	.192	.247	.130	.114	.063	.215	-.067
NB SRU	.020	.063	.053	.019	.160	.009	.136	-.004
NB NSRU	.546	.572	.437	.350	.163	.162	.365	-.119
Que SRU	.053	.044	.066	.053	.112	.023	.048	.017
Que NSRU	.336	.217	.415	.246	.534	.055	-.058	-.060
Ont SRU	.067	.060	.146	.076	.225	.071	.075	.022
Ont NSRU	.319	.005	.541	.017	.431	.056	.296	.005
Man SRU	-.014	.024	.069	.002	.117	-.003	.074	.088
Man NSRU	.154	.023	.019	.096	.308	.096	-.214	.313
Sask SRU	-.011	.014	.032	.048	.022	.017	.034	.054
Sask NSRU	.086	-.022	.018	.282	.077	.208	-.097	.139
Alta SRU	-.024	.016	.058	-.001	.041	.024	.006	.046
Alta NSRU	-.044	.144	-.013	.050	.463	.292	.041	-.028
BC SRU	.047	.046	.018	.063	.167	.004	.078	.066
BC NSRU	.114	.018	.507	-.106	-.016	.026	.225	.029
Can SRU	.055	.050	.102	.063	.173	.045	.065	.028
Can NSRU	.287	.129	.437	.135	.429	.093	.142	.000

TABLE III: $\hat{I}_{p2}$ and $\hat{I}_{pT1}$, respectively, by province NSRU and $\bar{I}_2$ and $\bar{I}_{T1}$ respectively for Canada NSRU (Mar-Dec 1975 Averages)

Sub-pop	Emp	Unemp	Emp Ag	Non-Ag	Manuf	Constr	TPU	Trade
				Characteristic				
Nfld	.175	.215	-.063	.105	.167	.124	-.090	.012
	.212	.177	.016	.180	.244	.351	-.068	-.081
PEI	.250	.392	-.152	.000	.327	.071	.061	-.229
	.250	.392	-.152	.000	.327	.071	.061	-.229
NS	.174	.033	.019	.103	.160	.026	.111	.015
	.354	.192	.247	.130	.114	.063	.215	-.067
NB	.355	.283	.149	.449	.164	.191	.300	.053
	.546	.572	.437	.350	.163	.162	.365	-.119
Que	.152	.103	.110	.122	.318	.088	-.071	.156
	.336	.217	.247	.246	.534	.055	-.058	-.060
Ont	.243	.002	.325	.224	.324	-.001	.110	.045
	.319	.005	.541	.017	.431	.056	.296	.005
Man	.196	-.008	.086	.109	-.077	-.055	-.007	.115
	.154	.023	.019	.096	.308	.096	-.214	.313
Sask	.134	.044	.097	.377	.085	.290	-.108	.271
	.086	-.022	.018	.282	.077	.208	-.097	.139
Alta	-.079	.045	.059	.041	.275	.198	.031	.142
	-.044	.144	-.013	.050	.463	.292	.041	-.028
BC	.148	.063	.400	-.062	-.052	.213	-.014	.006
	.114	.018	.507	-.106	-.016	.026	-.225	.029
Can	.181	.064	.224	.168	.265	.095	.025	.104
	.287	.129	.436	.135	.429	.093	.142	.000

TABLE IV: $\hat{I}_{M3}$ by Met Areas and $\hat{\hat{I}}_3$ Over 10 Met Areas, By Characteristic With $\hat{\hat{I}}_{T1}$ For Comparison (Mar-Dec, 1975 Averages)

Met Area	Emp	Unemp	Emp Ag	Non-Ag	Manuf	Constr	TPU	Trade
				Characteristic				
Halifax	.050	-.028	.140	.064	.058	.090	.047	.047
Quebec City	.179	.169	.234	.301	.269	.049	.100	.000
Montreal	.014	.060	.070	.032	.031	.028	.049	.031
Ottawa	.017	.108	.238	.033	.315	-.023	-.032	.002
Toronto	.048	.065	.209	.057	.108	.156	.071	.016
Hamilton	.003	.028	.085	-.002	.150	.036	-.046	.047
Winnipeg	.022	.029	.166	.038	.129	-.029	.061	.073
Calgary	-.040	.039	.204	-.001	.049	.053	-.028	.067
Edmonton	.009	.011	.035	.031	.041	.017	.003	.029
Vancouver	.028	.034	.020	.047	.063	.006	.020	.077
Can SRU $\hat{\hat{I}}_3$	.036	.064	.142	.059	.107	.069	.045	.031
Can SRU $\hat{\hat{I}}_{T1}$	.055	.050	.102	.063	.173	.045	.065	.028

FORMULAS FOR CORRELATION COEFFICIENTS UNDER SPECIAL RELATIONSHIPS BETWEEN ESTIMATED CHARACTERISTICS

Richard Griffin and Gary M. Shapiro
U. S. Bureau of the Census

I. INTRODUCTION

The following are some examples of situations in which the results of this article are applicable.

A. Suppose you have conducted a survey in which double sampling was used and you are interested in the ratio of Spanish persons with less than 12 years of education to total Spanish persons, where one quantity comes from the large sample and one from the small sample. This article gives a simple algebraic formula for the correlation coefficient between them so that the variance of the ratio can be estimated.

B. Suppose you want to compare the unemployment rate for Blacks to that of the total population. This article gives a simple algebraic formula for the correlation coefficient between the two estimated rates and also for the variance of the difference between the two rates.

The major impetus for this article came from a desire to estimate covariances among estimates from different subsamples of the 1970 Decennial Census. This was part of an effort to determine standard errors for the ratio of the number of certain minority groups who are U.S. citizens 18 years of age and older to the total number of U.S. citizens 18 years of age and older in individual counties. [2] This was of interest to us in connection with determining which jurisdictions should be included in coverage under the 1975 amendments to the Voting Rights Act. Consider, for example, the minority group Spanish.

Data on whether a person is of Spanish heritage or not is available from the 15 percent subsample of the decennial census. Citizenship data, however, is available from the 5 percent subsample. Thus we had to deal with ratios involving data from different subsamples of the census. We used Taylor series approximations on the ratios which left us with sums involving variances and covariances. Estimates of the variances were available. Many of the formulas in this article were derived for use in estimating the covariances.

Section II of this article gives the basic notation and assumptions. Section III contains all the formulas for correlation coefficients and, in four cases, the related variance estimates. Finally, the appendix covers proofs of all results given in section III.

It should be noted that the results in this paper have been proven only for simple random sampling. However, the authors have applied these results to systematic cluster samples and believe they yield reasonable approximations.

II. NOTATION

Suppose we have a population of N units. A simple random sample without replacement (S.R.S.W.O.R.) of size n_α is taken from this population. From the units in this sample, a S.R.S.W.O.R. of size n_β is taken. Among the n_α units in the first sample, there are n_γ units not selected in the second sample ($n_\alpha = n_\gamma + n_\beta$). The second sample is also a S.R.S.W.O.R. of size n_β from the entire population of N units. It is also true that the n_γ remaining units constitute a S.R.S.W.O.R. of size n_γ from the entire population.

Let x_A, x_B, and x_C be sample estimates of the number of units in the population that have characteristics A, B, and C respectively. Estimate x_A is calculated from the first sample, estimate x_B from the second sample, and estimate x_C from the n_γ remaining units. X_A, X_B, and X_C are the respective expected values of x_A, x_B, x_C.

Let M_A, M_B, and M_C represent the sets of all units in the population that have characteristics A, B, and C respectively. Also let $f_\alpha = \frac{n_\alpha}{N}$, $f_\beta = \frac{n_\beta}{N}$, and $f_\gamma = \frac{n_\gamma}{N}$. V is used to denote the coefficient of variation.

Let $x_{A'}$, $x_{B'}$, $M_{A'}$, $M_{B'}$, $X_{A'}$, $X_{B'}$ be defined in the same manner as x_A, x_B, M_A, M_B, X_A, and X_B respectively, except that A' and B' are different characteristics than A and B.

III. RESULTS

A. Let $M_A \subset M_B$, then $\rho_{x_A, x_B} =$

$$\frac{(1-f_\alpha) f_\beta}{f_\alpha (1-f_\beta)} \frac{V_{x_B}}{V_{x_A}}$$

This formula applies when the set of all units in the population that have the characteristic estimated from the large sample is a subset of the set of units in the population that have the characteristic estimated from the subsample. For example, consider unemployment and civilian labor force where civilian labor force is estimated from a subsample of the sample from which unemployment is estimated.

1. Special Case #1:

$$n_\alpha = n_\beta \Rightarrow \rho_{x_A, x_B} = \frac{V_{x_B}}{V_{x_A}}$$

This formula applies in the case where the set of units that have one of the characteristics is a subset of the set of units that have the other characteristic and both characteristics are estimated from the same sample.

2. Special Case #2:

$$A{=}B \Rightarrow \rho_{x_A, x_B} = \sqrt{\frac{\frac{n_\beta}{n_\alpha} - f_\beta}{1-f_\beta}}$$

(This result due to Bershad [1].)

This case applies when there is one characteristic estimated twice, the second estimate being formulated from a subsample of the sample used to produce the first estimate.

B. Let $M_B \supseteq M_A$, then $\rho_{x_A, x_B} = \frac{V_{x_A}}{V_{x_B}}$

This formula applies when the set of units in the population that have the characteristic estimated from the subsample is a subset of the set of units in the population that have the characteristic estimated from the large sample.

C. Let $M_C \subseteq M_B$, then $\rho_{x_B, x_C} = \frac{-f_\beta}{1-f_\beta} \frac{V_{x_B}}{V_{x_C}}$

This formula applies when a sample is taken and one characteristic is estimated from a subsample of the sample and the other characteristic is estimated from the units of the first sample that are not in the subsample. Furthermore, the set of units that have one of the characteristics must be a subset of the set of units that have the other characteristic. For example, consider lawyers and white collar workers where the number of lawyers is estimated from the 1970 Census 5 percent sample and the number of white collar workers is estimated from the 1970 Census 15 percent sample.

1. Special Case:

$$B{=}C \Rightarrow \rho_{x_B, x_C} = -f_\alpha \frac{\sqrt{\frac{n_\beta}{n_\alpha}\frac{n_\gamma}{n_\alpha}}}{\sqrt{(1-f_\beta)(1-f_\gamma)}}$$

(This result due to Bershad [1].)

This case applies when there is one characteristic estimated twice, first from a subsample of an original sample and secondly from the units of the original sample that are not in the subsample.

D. Let $M_A \cap M_B = \phi$,

then $\rho_{x_A, x_B} = -\frac{1-f_\alpha}{f_\alpha(N-1)} \frac{1}{V_{x_A} V_{x_B}}$

$$VAR(x_A - x_B) = \sigma^2_{x_A} + \sigma^2_{x_B} + \frac{2\, X_A\, X_B(1-f_\alpha)}{f_\alpha(N-1)}$$

This is applicable when the set of units that have the characteristic estimated from the large sample is disjoint from the set of units that have the characteristic estimated from the subsample.

E. Let $M_B \cap M_C = \phi$,

then $\rho_{x_B x_C} = \frac{1}{(N-1)V_{x_B} V_{x_C}}$

$$VAR(x_B - x_C) = \sigma^2_{x_B} + \sigma^2_{x_C} - \frac{2\, X_B X_C}{(N-1)}$$

This applies when a sample is taken and one characteristic is estimated from a subsample of the sample and the other characteristic is estimated from the units in the first sample that are not in the subsample. Furthermore, the set of units that have one of the characteristics is disjoint from the set of units that have the other characteristic. For example, consider college graduates and persons with only a high school education where the number of college graduates is estimated from the 1970 Census 15 percent sample and the number of persons with only a high school education is estimated from the 1970 Census 5 percent sample.

F. Let $M_B \subseteq M_A$, $M_{B'} \subseteq M_B$, $M_{A'} \subseteq M_A$, $M_{B'} \subseteq M_{A'}$,

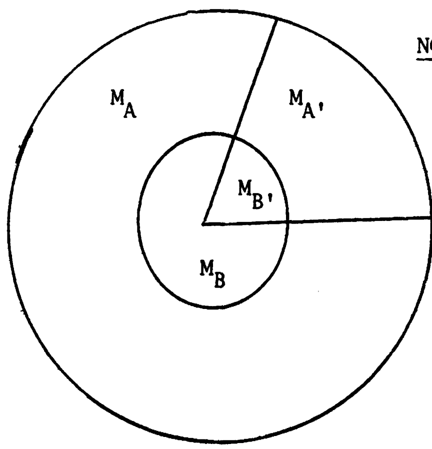
$n_\alpha = n_\beta$, $p_{A'} = \frac{x_{A'}}{x_A}$, and $p_{B'} = \frac{x_{B'}}{x_B}$,

then $\rho_{p_{A'}, p_{B'}} \cong \frac{X_B \sigma_{p_{B'}}}{X_A \sigma_{p_{A'}}}$ and

$$VAR(p_{A'} - p_{B'}) \cong \sigma^2_{p_{A'}} + \left(1 - 2\frac{X_B}{X_A}\right)\sigma^2_{p_{B'}}$$

(This result due to Tomlin [3].)

Illustration:



NOTE: M_A is the whole circle and M_B is the whole inner circle.

As an example of a situation in which these formulas apply, consider the unemployment rate for blacks and the overall unemployment rate, where both are estimated from the same sample.

G. Let $A=B$, $M_{A'} \subset M_A$, $M_{B'} \subset M_B$, $M_{A'} \cap M_{B'} = \phi$,

$n_\alpha = n_\beta = n$, $p_{A'} = \frac{x_{A'}}{x_A}$, $p_{B'} = \frac{x_{B'}}{x_B}$, $P_{A'} = \frac{X_{A'}}{X_A}$,

and $P_{B'} = \frac{X_{B'}}{X_{B'}}$

then

$$\rho_{p_{A'},p_{B'}} \cong -\sqrt{\frac{P_{A'}P_{B'}}{(1-P_{A'})(1-P_{B'})}} \quad \text{and}$$

$$\mathrm{VAR}(p_{A'} - p_{B'}) \cong \sigma^2_{p_{A'}} + \left(1 + \frac{2P_{A'}}{1-P_{B'}}\right) \sigma^2_{p_{B'}}$$

(This result due to Tomlin [3].)

As an example of a situation in which these formulas apply, consider the proportion of employed persons who are chemists and the proportion of employed persons who are engineers, where both are estimated from the same sample.

IV. ACKNOWLEDGEMENTS

We would like to thank Paul H. Tomlin and Kirk Wolter, both of the Bureau of the Census, for helpful comments on this paper.

APPENDIX - PROOFS OF RESULTS

I. Proof of A

$$\mathrm{cov}(x_A, x_B) = \mathrm{cov}(E(x_A/n_\alpha), E(x_B/n_\alpha)) + E[\mathrm{cov}(x_A, x_B/n_\alpha)]$$

$$= \mathrm{cov}(x_A, y_B) + 0$$

where $y_B = \frac{N}{n_\alpha} \sum_{i=1}^{n_\alpha} s_i$, where $s_i = \begin{cases} 1 & \text{if unit } i \text{ has characteristic B} \\ 0 & \text{otherwise} \end{cases}$

$$\left(\text{since } E\left(\frac{N}{n_\beta} \sum_{i=1}^{n_\beta} s_i/n_\alpha\right) = y_B\right)$$

now $\mathrm{cov}(x_A, y_B) = N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} \left(\sum_{i=1}^{N} r_i s_i - \frac{\sum_{i=1}^{N} r_i \sum_{i=1}^{N} s_i}{N} \right)$ (r_i is the indicator function for characteristic A)

$$= N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} X_A \left(1 - \frac{X_B}{N}\right)$$

thus
$$\rho_{x_A, x_B} = \frac{N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} X_A \left(1 - \frac{X_B}{N}\right)}{\sigma_{x_A} \sigma_{x_B}}$$

$$= \frac{(N-n_\alpha)n_\beta}{n_\alpha(N-n_\beta)} \frac{V_{x_B}}{V_{x_A}} \left(\text{since } \sigma^2_{x_B} = N \frac{N-n_\beta}{n_\beta} \frac{1}{N-1} X_B \left(1 - \frac{X_B}{N}\right)\right) = \frac{(1-f_\alpha)f_\beta}{f_\alpha(1-f_\beta)} \frac{V_{x_B}}{V_{x_A}}$$

II. Proof of B

Using the same argument as in A, we have:

$$\rho_{x_A, x_B} = \frac{N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} X_B \left(1 - \frac{X_A}{N}\right)}{\sigma_{x_A} \sigma_{x_B}}$$

$$= \frac{V_{x_A}}{V_{x_B}}$$

III. Proof of C

$$\mathrm{cov}(x_B, x_C) = \mathrm{cov}(E(x_B/n_\alpha), E(x_C/n_\alpha)) + E(\mathrm{cov}(x_B, x_C/n_\alpha))$$

Consider

$$E(x_B x_C/n_\alpha) = E\left(\frac{N \sum_{i=1}^{n_\alpha} v_i s_i}{n_\beta} \; \frac{N \sum_{i=1}^{n_\alpha} w_i t_i}{n_\gamma} \right)$$

(s and t are indicator functions for characteristics B and C respectively and v and w are the indicator functions for the samples of size n_β and n_γ, respectively.)

$$= \frac{N^2}{n_\beta n_\gamma} \left(\sum_{i=1}^{n_\alpha} s_i t_i \, E(v_i w_i) + \sum_{i \neq j}^{n_\alpha} s_i t_j \, E(v_i w_j) \right) = \frac{N^2}{n_\beta n_\gamma} \frac{n_\beta}{n_\alpha} \frac{n_\gamma}{(n_\alpha - 1)} \sum_{i=1}^{n_\alpha} t_i \left(\sum_{i=1}^{n_\alpha} s_i - 1 \right)$$

so

$$E(\mathrm{cov}(x_B x_C/n_\alpha)) = E\left[\frac{N^2}{n_\alpha(n_\alpha - 1)} \sum_{i=1}^{n_\alpha} t_i \left(\sum_{i=1}^{n_\alpha} s_i - 1 \right) - \frac{N^2}{n_\alpha^2} \sum_{i=1}^{n_\alpha} s_i \sum_{i=1}^{n_\alpha} t_i \right]$$

$$= \frac{N^2}{n_\alpha^2 (n_\alpha - 1)} \left(\sum_{i=1}^{N} t_i s_i \, E(u_i^2) + \sum_{i \neq j}^{N} t_i s_j \, E(u_i u_j) \right) - \frac{N^2}{n_\alpha (n_\alpha - 1)} \sum_{i=1}^{N} t_i \, E(u_i)$$

(u is the indicator function for the sample of size n_α.)

$$= \frac{N^2}{n_\alpha^2 (n_\alpha - 1)} \left(X_C \frac{n_\alpha}{N} + X_C (X_B - 1) \frac{n_\alpha (n_\alpha - 1)}{N(N-1)} \right) - \frac{N^2}{n_\alpha (n_\alpha - 1)} X_C \frac{n_\alpha}{N}$$

$$= \frac{N}{n_\alpha (n_\alpha - 1)} X_C + \frac{N}{n_\alpha (N-1)} X_C (X_B - 1) - \frac{N}{(n_\alpha - 1)} X_C$$

$$= \frac{N - N n_\alpha}{n_\alpha (n_\alpha - 1)} X_C + \frac{N}{n_\alpha (N-1)} X_C (X_B - 1)$$

next

$$\mathrm{cov}(E(x_B/n_\alpha), E(y_C/n_\alpha)) = N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} X_C \left(1 - \frac{X_B}{N}\right)$$

and then

$$\rho_{x_B, x_C} = \frac{N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} X_C \left(1 - \frac{X_B}{N}\right) + \frac{N - N n_\alpha}{n_\alpha (n_\alpha - 1)} X_C + \frac{N}{n_\alpha (N-1)} X_C (X_B - 1)}{\sigma_{x_B} \sigma_{x_C}} = \frac{-f_B}{1 - f_B} \frac{V_{x_B}}{V_{x_C}}$$

(This equality results after the use of a lot of algebra which is not presented here.)

IV. Proof of D

Using arguments similar to those used in A and B, we have:

$$\rho_{x_A, x_B} = \frac{-N \frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} \frac{X_A X_B}{N}}{\sigma_{x_A} \sigma_{x_B}} = -\frac{1-f_\alpha}{f_\alpha (N-1)} \frac{1}{V_{x_A} V_{x_B}}$$

V. Proof of E

$$\mathrm{cov}(x_B, x_C) = E\left[\mathrm{cov}(x_B, x_C/n_\alpha)\right] + \mathrm{cov}\left[E(x_B/n_\alpha),\ E(x_C/n_\alpha)\right]$$

now $\mathrm{cov}\left(E(x_B/n_\alpha),\ E(x_C/n_\alpha)\right) = \mathrm{cov}\ (y_B, y_C) = -\frac{N-n_\alpha}{n_\alpha} \frac{1}{N-1} X_B X_C$

(y_B is defined as in the proof of A and y_C is defined in the same manner.)

Next, consider

$$E\left[\mathrm{cov}(x_B, x_C/n_\alpha)\right] = E\left[E(x_B x_C/n_\alpha) - E(x_B/n_\alpha)\ E(x_C/n_\alpha)\right]$$

$$E(x_B x_C/n_\alpha) = E\left(N \frac{\sum_{i=1}^{n_\alpha} v_i s_i}{n_\beta} \frac{N \sum_{i=1}^{n_\alpha} w_i t_i}{n_\gamma}\right)$$

(s and t are the indicator functions for characteristics B and C respectively and v and w are the indicator functions for the sample of size n_β and the sample of size of n_γ respectively.)

$$= \frac{N^2}{n_\beta n_\gamma} \left(\sum_{i=1}^{n_\alpha} s_i t_i\ E(v_i w_i) + \sum_{i \neq j}^{n_\alpha} s_i t_j\ E(v_i w_j)\right)$$

$$= \frac{N^2}{n_\beta n_\gamma} \sum_{i \neq j}^{n_\alpha} s_i t_j\ E(v_i w_j) = \frac{N^2}{n_\beta n_\gamma} \frac{n_\beta}{n_\alpha} \frac{n_\gamma}{n_\alpha - 1} \sum_{i \neq j}^{n_\alpha} s_i t_j$$

and

$$E(x_B/n_\alpha) \quad E(x_C/n_\alpha) = \frac{N^2}{n_\alpha^2} \sum_{i=1}^{n_\alpha} s_i \sum_{i=1}^{n_\alpha} t_i = \frac{N^2}{n_\alpha^2} \sum_{i \neq j}^{n_\alpha} s_i t_j$$

so

$$E\left[\mathrm{cov}(x_B, x_C/n_\alpha)\right] = E\left(\frac{N^2}{n_\alpha(n_\alpha - 1)} \sum_{i \neq j}^{n_\alpha} s_i t_j - \frac{N^2}{n_\alpha^2} \sum_{i \neq j}^{n_\alpha} s_i t_j\right)$$

$$= \frac{N^2}{n_\alpha^2 (n_\alpha - 1)} E\left(\sum_{i \neq j}^{n_\alpha} s_i t_j\right) = \frac{N^2}{n_\alpha^2 (n_\alpha - 1)} \sum_{i \neq j}^{N} s_i t_j\ E(u_i u_j)$$

(u is the indicator function for the sample of size n_α.)

$$= \frac{N^2}{n_\alpha^2 (n_\alpha - 1)} X_B X_C \frac{n_\alpha (n_\alpha - 1)}{N(N-1)} = \frac{N}{n_\alpha (N-1)} X_B X_C$$

Thus

$$\rho_{x_B, x_C} = \frac{\frac{N}{n_\alpha (N-1)} X_B X_C - \frac{N-n_\alpha}{n_\alpha} \frac{1}{(N-1)} X_B X_C}{\sigma_{x_B} \sigma_{x_C}} = \frac{\frac{1}{N-1} X_B X_C}{\sigma_{x_B} \sigma_{x_C}}$$

VI. Proof of F

Let $M_D = M_A \cap M_B^C$ and $P_{D'}$ be the estimated proportion of units in M_D that have characteristic A'. (M_B^C is the complement of M_B). $\mathrm{cov}(p_{B'}, p_{D'}) = E\left[\mathrm{cov}(p_{B'}, p_{D'})/(x_B, x_D)\right]$

$+ \text{cov}\left[E\,(p_{B'}/(x_B, x_D)),\ E(p_{D'}/(x_B, x_D))\right] = 0$, where x_D is the sample estimate of the number of units in the population that have characteristic D. Note that for cluster sampling rather than simple random sampling, this covariance may not always be close to zero.

Now, $p_{A'} \cong \frac{X_B}{X_A} p_{B'} + \frac{X_D}{X_A} p_{D'}$ Thus, $\text{cov}(p_{A'}, p_{B'}) \cong \frac{X_B}{X_A} \text{VAR}(p_{B'}) + \frac{X_D}{X_A} \text{cov}\,(p_{B'}, p_{D'}) = \frac{X_B}{X_A} \text{VAR}(p_{B'})$

$$\Rightarrow \rho_{p_{A'}, p_{B'}} \cong \frac{X_B \sigma_{p_{B'}}}{X_A \sigma_{p_{A'}}}$$

So that $\text{VAR}(p_{A'} - p_{B'}) \cong \sigma^2_{p_{A'}} + \sigma^2_{p_{B'}} - 2\frac{X_B}{X_A}\sigma^2_{p_{B'}} = \sigma^2_{p_{A'}} + \left(1 - 2\frac{X_B}{X_A}\right)\sigma^2_{p_{B'}}$

VIII. Proof of G

Let r, r', and s' be the indicator functions for characteristics A, A', and B' respectively. Let u be the indicator function for the sample.

Then let $k = \sum_{i=1}^{N} u_i r_i$, $\ell = \sum_{i=1}^{N} u_i r'_i$, and $m = \sum_{i=1}^{N} u_i s'_i$

So $\text{cov}\,(p_{A'}, p_{B'}) = \text{cov}\left(\frac{\ell}{k}, \frac{m}{k}\right) = \text{cov}(E(\frac{\ell}{k}/k),\ E(\frac{m}{K}/k)) + E(\text{cov}(\frac{\ell}{k}, \frac{m}{k}/k)) = 0 + E\ (\frac{1}{K^2}\ \text{cov}(\ell, m/k))$

$E(\ell/k) = k\ P_{A'}$ and $E(m/k) = k\ P_{B'}$ $\quad E(\ell m/k) = E\left(\sum_{i=1}^{X_A} z_i r'_i \sum_{i=1}^{X_A} z_i s'_i\right)$

where z is the indicator function for a S.R.S.W.O.R. of k units that have characteristic A from the population of units that have characteristic A.

$$= E\left(\sum_{i=1}^{X_A} z_i^2 r'_i s'_i + \sum_{i \neq j}^{X_A} z_i z_j r'_i s'_j\right) = \sum_{i \neq j}^{X_A} r'_i\ s'_j\ E(z_i z_j) = \frac{k(k-1)}{X_A(X_A-1)} X_{A'} X_{B'}$$

Thus $\text{cov}(\ell, m/k) = \frac{k(k-1)}{X_A(X_A-1)} X_{A'} X_{B'} - \frac{k^2 X_{A'}}{X_A}\frac{X_{B'}}{X_B} = -k\ P_{A'} P_{B'} \left(k - \frac{X_A}{X_A - 1}(k-1)\right)$ (since A=B) $\cong -k\ P_{A'} P_{B'}$

(A good approximation for large X_A)

So, $\text{cov}(p_{A'}, p_{B'}) \cong E\left(-\frac{1}{k} P_{A'} P_{B'}\right) \cong -\frac{1}{k} P_{A'} P_{B'}$

From this we have, $\text{VAR}(p_{A'} - p_{B'}) \cong \frac{P_{A'}(1-P_{A'})}{k} + \frac{P_{B'}(1-P_{B'})}{k} + 2\frac{P_{A'}P_{B'}}{k} \cong \sigma^2_{p_{A'}} + \sigma^2_{p_{B'}}\left(1 + \frac{2P_{A'}}{1-P_{B'}}\right)$

and,

$$\rho_{p_{A'}, p_{B'}} \cong -\sqrt{\frac{P_{A'}P_{B'}}{(1-P_{A'})(1-P_{B'})}}$$

REFERENCES

1. Bershad, Max A., "Correlations Between Half Samples," Internal Census Bureau memorandum to William N. Hurwitz, July 1, 1968.

2. Miskura, Susan, "Documentation of the Derivation of Estimates and Variance Estimates for the Voting Rights Amendment Tabulations," Internal Census Bureau memorandum to Gary M. Shapiro, November 10, 1975.

3. Tomlin, Paul H., "A Possible Method of Presentation of Comparisons of Proportions in Source and Reliability Statements," Internal Census Bureau memorandum to Gary M. Shapiro, February 19, 1976.

VALIDATION OF THEORIZED FACTOR STRUCTURE OF THE PERSONAL ORIENTATION INVENTORY

Carolyn C. Hamlett, Northeast Louisiana University

The work of Barker and Barker has produced a method which can be used to evaluate theories regarding factor structure of large data matrices. (Barker, B. M., & Barker, H. R., 1975; Barker, H. R., & Barker, B. M., 1975A, 1975B, 1976; Hunnicutt & Barker, 1974). The method employs an indirect factor analysis program developed by Barker (1973) which is based on a mathematical technique offered by Horst (1965). The procedure provides a way of performing a full-scale factor analysis of large data sets under specific conditions (Barker & Barker, 1976). Previously, such factor analyses have been rendered impractical or impossible due to several limitations. The first and most prohibitive of these limitations is that most computers can factor no more than one hundred variables at a time. Even when a greater computer capacity exists, the cost in computer time of factoring a large matrix can be prohibitive (Barker, Fowler, & Peterson, 1971).

The method proposed by Barker and Barker (1976) offers a means of factoring large data sets under specific conditions. Five years of research on Horst's indirect factor method summarized by Barker and Barker (1975) resulted in virtually perfect replication of a factor analysis using conventional factor analytic methods. This indirect factor analysis was accomplished in four minutes of computer time as opposed to 50 hours for the conventional solution. The purpose of this study was to demonstrate the value of the indirect method of factor analysis as a technique for testing hypotheses regarding factor structure of large data sets. Consequently, several theories regarding the Personal Orientation Inventory (POI) were selected from the literature for evaluation using the indirect method.

The POI, developed by Everett L. Shostrom, is frequently used to measure dimensions of personality. The POI consists of 150 pairs of value statements which are scored twice to yield scale scores purporting to measure aspects of self-actualization and personality orientation. Of these scales, the Time (TC) and Support (I) scales are expressed in terms of ratios. The other ten subscales yield a profile of the individual which is interpreted by comparison to that of a self-actualized person.

Various theories exist regarding dimensions of the POI. Shostrom himself does not conceptualize the POI scales "as representing independent dimensions" (1972, p. 21), and, in fact, several studies have indicated that subscales are intercorrelated (Damm, 1969; Klavetter & Mogar, 1967). Klavetter and Mogar (1967) hypothesize that the subscales may "lack unique variance" (p. 424), since the TC, I, and SAV scales were found to account for most of the test's variance. As a result of this study, Klavetter and Mogar conclude that twelve subscales are redundant and that a more accurate instrument would consist of fewer scales with less item overlap. In 1969 Damm conducted a study which supports the findings of Klavetter and Mogar. His findings indicate that the most useful subscales are the I and TC scales.

While these studies appear to indicate that the POI may have only two dimensions, another study, conducted in 1974 by Silverstein and Fisher, presents five item clusters comprised of 74 POI items. These five clusters were obtained by applying elementary linkage analysis to POI data and are free of item overlap.

In summary, evidence exists which appears to indicate that the POI may be measuring somewhat global aspects of the construct rather than twelve specific characteristics. The degree of intercorrelation which apparently exists between subscales appears to lend support to the hypothesis that the POI has fewer than twelve dimensions.

Methodology

Subjects of this study were students enrolled in the Colleges of Education, Home Economics, Engineering, and Arts and Sciences at the University of Alabama during the spring semester of the 1976 academic year. A sample of 501 subjects was obtained.

The 501 subject data base was factor analyzed by a conventional principal axis solution in order to generate an empirical set of findings to be compared with current theories. A conventional factor analysis program (CORR12) developed by Barker (1973) was used. None of the existing statistical programs at the University of Alabama computer installation could handle a matrix of this size (150 x 150), due to the enormous amount of core space required and to impractical length of run time. Consequently, the size of CORR12 was incremented so that this factor analysis could be carried out. CORR12 was developed in such a way that only the upper triangular matrix of correlations is worked with in the computer. This alone saves enormous computer core space and cuts computer run time drastically. Despite use of this procedure, at the present time CORR12 takes up almost all the allotted core space. Results of this conventional factor analysis constituted one of the theories to be tested regarding dimensions of the POI and provided a means of testing the value of the indirect factor method itself.

Other theories regarding dimensions of the POI selected for study were tested using the indirect method (CORR99) developed by Barker (1973). In addition to the theory based on a direct factor analysis, two theories from the literature were examined. The first of these comprised scales proposed by Shostrom (1972); the second, item clusters developed by Silverstein and Fisher (1974). A third theory was generated by randomly assigning variables to arbitrarily selected numbers of totals.

The information measure D (relative uncertainty reduction) was selected to provide an objective measure of the degree to which theorized dimensions reflect actual factor structure of the data set. In an ideal solution all cell entries in the matrix of factors would appear in the diagonal. Such a solution would indicate complete agreement of item subsets (totals) with actual factor structure. Frequently, however, certain items are found to load inappropriately. Items which fail to load as expected onto factors appear as false negatives. Those which load onto factors contrary to expectation appear as false positives. The D measure expresses the relationship between rows (subsets of items) and columns (factors). Use of this statistic also permits comparison between theories of degree of agreement between a priori item subsets and actual factor structure.

Results

Theories tested were rank ordered with respect to adequacy (most adequate = 1.0) in reflecting factor structure of the POI as follows:

1. Five factors based on results of direct factor analysis of the data set, D = .94.
2. Ten subscales hypothesized by Shostrom, D = .79.
3. Five item clusters proposed by Silverstein and Fisher, D = .52.
4. Shostrom's ratio scales, I and TC; D = .41.
5. Randomly formed totals, D = .41.

The relatively low D measures obtained for Shostrom's theory (Tables 2 and 4) and the theory of Silverstein and Fisher (Table 3) may be due to several factors. Both Shostrom's subscales and the item clusters proposed by Silverstein and Fisher do not include all 150 POI items. It was to be expected that different configurations of item loadings would result from factor analysis of all 150 variables and, while factorial purity of Shostrom's subscales was tested using only 54 items due to omission of overlapping items, it is likely that had Shostrom's ten subscales been tested intact, without taking into account item overlap, a different solution would have been obtained.

Other variables to be considered include the fact that Shostrom does not hypothesize the 12 POI scales as reflecting factorially pure dimensions. Construction of the POI did not include factor analysis of items. Silverstein and Fisher (1974) describe problems encountered in factor analysis of the POI which resulted in their turning to elementary linkage analysis in order to examine structure of the POI. The relatively low D measures obtained when their proposed item clusters were evaluated may reflect the inadequacy of the method employed in forming these five clusters or differences inherent in samples used. (Silverstein and Fisher used a sample from a prison population; the present study, a sample of supposedly normal subjects.)

In order to generate a third theory to be tested, items comprising the POI were factored c conventionally using a principal axis solution. When the scree test (Cattell, 1966) was applied to the eigenvalues, five factors seemed important. These five factors were rotated to a varimax criterion. An item was identified with a rotated factor if its varimax load was equal to or exceeded + or - .3. Items whose loads equaled or exceeded + or - .3 on more than one factor were omitted from the analysis.

The varimax rotation of the five factor solution resulted in identification of 84 items. It was noted that a surprisingly large number of items (63) failed to load onto any of these five factors. Furthermore, in most cases item loads used in identifying items with factors were found to range between + or - .3 and + or - .5. Based on identification of items with factors, the five factors were named as follows: (1) Self-perception; (2) Self-actualizing Values; (3) Existentiality; (4) Sensation Responsiveness; and (5) Inner-other Support.

Direct factor analysis resulted in an impressive degree of agreement of factor structure and a priori item subsets (Table 5). Three factors were perfectly identified by the indirect method and a four factor resulted in only two false negatives and one false positive. The remaining factor resulted in seven false negatives which is disappointing. Nevertheless, the obtained D measure indicates a high degree of similarity between theorized and actual factor structure.

While the rive factor solution appears to be an accurate reflection of factor structure of the POI, examination of these results appears to indicate possible deficiencies in the POI itself. Factors which emerged from the direct factor analysis do not appear clean. The cumulative proportion of variance (.17) associated with the solutions is not impressive and the number of POI items (84) identified with factors was surprisingly small. Other items which failed to meet the criterion of loading + or - .3 or better on a factor presumably loaded on negligible factors. Furthermore, most items which met the criterion set for identification with a factor had loadings which ranged between + or - .3 and + or - .5. Such loads are low.

Previous studies (Damm, 1969; Klavetter & Mogar, 1967; Silverstein & Fisher, 1974) indicate that the POI would be more useful and actual dimensions more accurately expressed if it were made up of fewer than 12 scales. The present study supports this contention. These results indicate that the POI should be examined through means of item analysis, further factor analytic studies, and other suitable techniques in order to reduce the number of test items to those which appear to contribute to reliability and factorial purity of the scales.

Results of this study support earlier studies by Barker and Barker (1975B; 1976) in demonstrating validity of the indirect method as an appropriate technique in evaluating theories regarding factor structure of large data sets. The direct factor analysis represents a criterion by which accuracy of the indirect method can be measured. As expected, the obtained D measure of .94 indicates a high degree of agreement between the a priori item subsets based on the direct factor analysis and actual factor structure as measured by the indirect method.

Of additional interest are differences in cost and in computer time between applications of a conventional factor analysis and the indirect method. Two direct factor analyses of the POI required more than 25 minutes of computer run time and $147 charged to the computer account. Five analyses using the indirect method cost a total of $12.16 and approximately 12 minutes of computer run time. These figures illustrate that, under certain specific conditions, the indirect method has great practical value over more conventional factor analytic programs in analyses of very large data sets.

References

Barker, B. M., & Barker, H. R. Search for a method of increasing the accuracy of a novel, indirect method of factor analysis. Paper presented at Southeastern Psychological Association, Atlanta, March 1975.

Barker, H. R. Behavioral Sciences Statistical Program Library. University, Al.: University of Alabama Reproduction Services, 1973.

Barker, H. R., & Barker, B. M. Accuracy of an indirect factor analytic solution as a function of the number of totals used. Paper presented at Southeastern Psychological Association, Atlanta, March 1975. (A)

Barker, H. R., & Barker, B. M. Summary of research on a novel, indirect factor analytic solution. Proceedings of the American Statistical Association, Atlanta, August, 1975. (B)

Barker, H. R., & Barker, B. M. An indirect factor method for testing the dimensionality of large data sets. Paper presented at Southeastern Psychological Association, New Orleans, March 1976.

Barker, H. R., Fowler, R. D., & Peterson, L. P. Factor analytic structure of the short form MMPI items in a VA hospital population. Journal of Clinical Psychology, 1971, 27, 228-233.

Hunnicutt, B. M., & Barker, H. R. Application and evaluation of a novel and indirect factor method to a large data matrix. Paper presented at Southeastern Psychological Association, Hollywood, May 1974.

Cattell, R. B. (Ed.). Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally, 1966.

Damm, V. J. Overall measures of self-actualization derived from the Personal Orientation Inventory. Educational and Psychological Measurement, 1969, 29, 977-981.

Klavetter, R. E., & Mogar, R. E. Stability and internal consistancy of a measure of self-actualization. Psychological Reports, 1967, 21, 422-424.

Silverstein, A. B., & Fisher, G. Cluster analysis of Personal Orientation Inventory items in a prison sample. Multivariate Behavioral Research, 1974, 9, 325-330.

Shostrom, E. L. Manual for the Personal Orientation Inventory. San Diego, Calif.: Educational and Industrial Testing Service, 1972.

TABLE 1
Association Between Item Subsets and Varimax Factors
(Randomly Formed Totals)

	I	II	III	IV	V	false −	sum
1	4					26	30
2		11				19	30
3			3			27	30
4				5		25	30
5					8	22	30
false +							0
sum	4	11	3	5	8	119	150

$H_x = 1.18$ $H_{xy} = 3.01$

$H_y = 2.31$ $D = .41$

TABLE 2

Association Between Item Subsets and Varimax Factors
(Shostrom's Support and Time Ratios)

	I	II	false −	2
1	11		12	23
2		4	123	127
false +				0
	11	4	135	150

$H_x = .61$ $H_{xy} = .94$

$H_y = .55$ $D = .41$

TABLE 3

Association Between Item Subsets and Varimax Factors
(Theory Based on the Work of Silverstein and Fisher)

	I	II	III	IV	V	false −	sum
1	5					2	7
2		6				10	16
3			5			3	8
4				2		8	10
5					10	23	33
false +	1						1
sum	6	6	5	2	10	46	75

$H_x = 2.13$ $H_{xy} = 3.00$

$H_y = 1.80$ $D = .52$

TABLE 4

Association Between Item Subsets and Varimax Factors
(Shostrom's Ten Subscales)

	I	II	III	IV	V	VI	VII	VIII	IX	X	false −	sum
1	4											4
2		6									3	9
3			1									1
4				3							2	5
5					2							2
6						4					3	7
7							2					2
8								4			1	5
9									5			5
10										5	11	16
false +	3											3
sum	7	6	1	3	2	4	2	4	5	5	20	59

$H_x = 3.11$ $H_{xy} = 3.75$

$H_y = 3.01$ $D = .79$

TABLE 5

Association Between Item Subsets and Varimax Factors
(Theory Based on Direct Factor Analysis)

	I	II	III	IV	V	false −	sum
1	13					7	20
2		11					11
3			22			2	24
4				18			18
5					11		11
false +			1				1
sum	13	11	23	18	11	9	85

$H_x = 2.32$ $H_{xy} = 2.66$

$H_y = 2.51$ $D = .94$

AN EXPERIMENT TO COMPARE ALTERNATIVE VERIFICATION PROCEDURES FOR MORTALITY MEDICAL CODING

Kenneth W. Harris and Dwight K. French
National Center for Health Statistics

I. Background

One of the most difficult data processing jobs for the National Center for Health Statistics (NCHS) is the coding of medical conditions listed as causes of death in the annual file of almost two million death records. The medical portion of the certificate consists of three lines on which the attending physician or other official is instructed to enter the sequence of medical conditions that led to death, and another line for listing other significant conditions. A nosologist (medical coder) assigns numerical codes to the medical conditions according to the Eighth Revision of the International Classification of Diseases, Adapted for Use in the United States (ICDA). These codes serve as input to a computer program that assigns one condition, called the "underlying cause of death," to represent all conditions on a certificate.

The assignment of underlying causes, either by hand or by machine, is not subject to ongoing verification because the process of assignment is so accurate (less than one-half of one percent error) that a formal verification system is neither cost nor quality effective. However, the original condition codes, which are assigned by a large staff of coders with varying degrees of proficiency, are subject to sample verification. After the original (production) coder has completed a work lot, two other coders independently code a ten percent systematic sample of records. The two sets of sample records and the corresponding production records are matched by computer, line by line and position by position. If two coders have entered the same code in the same position on the same line of a record, that code is placed into a "correct" or "preferred" set of codes for the record. The third coder matches the code if she codes it anywhere on the same line, regardless of position; otherwise, she is charged with an error.

After the matching procedure is completed for an entire lot, estimates of lot error rates for all three coders are produced by dividing the number of errors charged to each coder by the sum of the numbers of preferred codes for the sample records. These error rates serve as input into the production standards system, which is used to evaluate coder performance. Also, the production coder's error rate (the estimate of the quality of the outgoing product) determines whether her work lot is acceptable for underlying cause processing. If her sample error rate is 5 percent or less, her work lot is accepted as she coded it; otherwise, the entire lot is recoded by a fourth distinct coder and rematched against the two original sample coders.

The three-way independent verification system for mortality medical coding was instituted in 1968 because it was considered a more reliable method of measuring the level of coding error in the data than the two-way dependent system previously in use. Studies on other types of data[1 2 3] have shown that independent verification yields truer estimates of the amount of error in the data than does dependent verification, because a dependent verifier is biased toward the work of the original coder. However, no thorough study has ever been conducted to test the accuracy of mortality medical coding error rates based on the three-way system.

This accuracy is open to question for the following reasons:

(1) Poor handwriting, incorrect or confusing placement of medical entities on the death certificate, or poor quality microfilm can make it impossible to unquestionably determine the correct code for one or more of the conditions on a certificate.

(2) Even if the certificate is clear and properly filled out, the coding instructions may be sufficiently vague to allow two or more acceptable codes for a particular entity. The appropriateness of the three-way independent system is based on the assumption that a medical condition leads to only one code, so that when two or three of three coders with comparable ability independently arrive at the same code, there is a high probability that the code is correct. If this assumption is not valid then a coder with an acceptable code will be charged with an error when the other two coders match codes.

(3) When two or three coders have nonmatching codes all three coders are charged with an error, although it is quite likely that at least one coder has an acceptable code.

The primary purpose of the experiment was to measure the accuracy of the error rates produced by three-way independent verification, and compare them with error rates produced by two other commonly used methods of verification: two-way dependent and two-way independent coding with adjudication of differences.

II. Design of the Experiment

A. Constraints and Their Effect on the Sample Size

The goals of the experiment, along with limitations on coding time and money available, imposed the following set of constraints on the design:

(1) It was necessary to conduct the experiment using death certificates that had already been processed by the three-way verification system.

(2) The number of work lots represented in the sample needed to be sufficiently large

to include both production and sample verification work for most of the coders on the staff.

(3) In order to compare the accuracy of error rates produced by different verification systems, some measure of the "true" amount of error in the data was needed. "Truth" would have to be determined by having a small group of "experts" code the data.

(4) The number of records in the sample needed to be small enough so that the "expert" coder would not be overburdened. Each expert was limited to a maximum of two weeks of coding time to finish her assignments.

(5) Relative standard errors (RSE's) of certain key estimates should be within fixed bounds.

The sampling frame for the experiment consisted of the 472 work lots of 1974 data that passed through the three-way verification system during the months of July 1974 through March 1975. From this universe, a sample of work lots was selected. The 10 percent quality control sample of certificates within each lot represented a second stage of sampling. The RSE constraint made it necessary to have at least 25 lots in the sample, and the constraint on expert coding time limited the sample size to about 30 lots. Therefore, the first stage sample size was set at 30 lots.

B. Selection of Sample Lots

Prior to selection of the first-stage sample, the 472 work lots were sorted into 10 production error rate strata. The lots in each stratum were ordered randomly, the strata were ordered from smallest error rate upward, and the lots were temporarily renumbered from 001 to 472. The sample lots were selected systematically, so that the first-stage was essentially equivalent to a proportionate stratified sample. This procedure assured that the sample lots would be representative of the 472 lots in terms of coding difficulty (and for that matter, representative of all mortality medical coding, because the content of death certificates changes very little from year to year). In addition, the sample lots encompassed a good cross section of the 1974 Mortality Medical Coding staff. The coders who had completed the most assignments during the data year were the ones represented most often in the experiment. All coders on the staff were represented at least once as a production coder or sample coder.

C. Coding Assignments

Within each sample lot the medical codes of six distinct coders for the 10 percent quality control sample served as input data for the experiment. The first three coders were the original production coder (hereafter referred to as coder 1), and the two independent sample coders (hereafter referred to as coders 2 and 3). Because these numerical designations correspond to the coder numbers on lot-by-lot quality control reports, it was possible to distinguish coders 2 and 3 for each lot. The original production coder was always coder 1, even if her work was rejected and the lot was recoded.

A fourth coder (referred to as coder 5) was assigned to code the sample records while having access to the work of the production coder. This assignment was intended to correspond to the dependent verification procedure that was used during the 1973 data year. Because dependent verification assignments for 1973 data were given to the best available coders, most of the verification was handled by coders with relatively low error rates. In order to follow this system as closely as possible, coder 5 assignments for the experiment were allocated to 10 of the 12 coders, exclusive of the top two, with the lowest average error rates for their work during the 1974 data year. Each of the 10 coders was given 3 randomly selected lots that she had not previously worked on as coder 1, 2, or 3.

A fifth coder, hereafter called E6, was assigned to code the sample records with the codes of the production coder and one of the sample coders available to her. Her role corresponded to that of a dependent adjudicator in a two-way independent verification system. If the work lot number was odd, E6 was given access to the work of coders 1 and 2; if it was even, she had access to the work of coders 1 and 3.

The coding instructions for E6 were somewhat different from the instructions for the previous four coders. Whereas the other coders entered only one set of codes to represent the causes of death listed on a certificate, E6 was instructed to list all sets of codes that she considered acceptable to represent the certificate. If she thought that each condition on a certificate had a single correct code, she entered one set of codes. However, if one or more conditions could be coded more than one way, she entered all possible acceptable sets, changing only the code(s) in question.

Whenever E6 considered more than one set of codes acceptable, she listed the sets based on the following rules:

(1) If two or more sets are acceptable, but one is clearly preferable, list that set first and write "P" next to that set.

(2) If two or more sets are equally acceptable, the first set listed is the one she would choose if she had to decide on one of the sets. Write "D" next to that set.

The set with the P or D code following it will be referred to in future discussion as the "set of first choice."

A final coder, hereafter called E7, was assigned to code the sample records without having access to anyone else's work. Her coding instructions were exactly the same as those of E6.

The work of E6 and E7 was treated as two measures of "truth" for the purposes of this experiment. It was important, therefore, that the coders assigned to these roles be the very best coders available.

In addition, the same group of coders had to be assigned the roles of E6 and E7 so that variation between coders would not cause variation between the work of the E6 and E7 groups. To satisfy these requirements, a group of 6 "experts" was designated for these two roles. Four of the six were supervisors of the Mortality Medical Coding Unit, and the other two were the top rated coders in the Unit. The E6 and E7 assignments were randomly distributed so that each expert had ten assignments, five as E6 and five as E7. No expert was given E6 and E7 assignments for the same lot, and no expert was given an assignment in a lot she had previously coded as a production coder, sample coder, or recoder. Since the average number of sample records per lot (300) was less than the minimum daily production standard for Mortality Medical Coders (425), each expert could be expected to finish the 10 assignments within 10 coding days (2 weeks), and thus satisfy the time constraint for expert coding.

III. Analysis

A. Record Match and Assignment of Errors

After coders 5, E6, and E7 completed their coding assignments, their codes were fed into a computer program along with the codes of the original three coders. The program matched each of the original coders (1, 2, 3) with each of the coders used in the experiment (5, E6, E7), assigning errors to the original coders when their codes did not match. Whenever E6 and/or E7 coded more than one set of codes, the program observed the following rules:

(1) The number of code comparisons (the denominator in computing the error rate) between each original coder and the expert was taken from the expert's set of first choice. However,

(2) The number of errors charged to each original coder was taken from the expert's set that minimized the number of errors. The number of codes in the expert's set of first choice was used so that the denominator for computing error rates would be the same for each original coder.

B. "Expert" Agreement

In order to compare the merits of various verification systems used in this experiment, it was first necessary to determine the "true" value of the statistic being estimated, i.e., the error rate of mortality production coding (coder 1). Since the E6 and E7 assignments were completed by the very best available nosologists, these two assignment groups provided us with the "true" error rate in the sample of thirty lots. With the recognition that the "truth" from two sources might not necessarily be the same (but hopefully would be very close), we measured the agreement between E6 and E7 for all cases where at least one of them entered a code. For records where E6 and/or E7 entered more than one set of codes, the following rules were established to determine which sets should be used to measure the agreement between them:

(1) Select the comparison that minimizes the number of differences between the two experts. If two or more comparisons yield the same number of differences,

(2) Select from that group the comparison that maximizes the number of agreements between the two experts. If two or more comparisons yield the same minimum number of differences and maximum number of agreements,

(3) Select one of those comparisons using the following priority order:

a) The comparison involving the set of first choice by both experts.

b) A comparison involving E7's set of first choice.

c) A comparison involving E6's set of first choice.

d) Any other comparison.

Whenever a difference between E6 and E7 occurred, the correct code was credited to one of the experts if at least three of the other coders (1, 2, 3, 5) matched her code. The other expert was charged with an error. If no expert had a 3-1 or 4-0 majority match, neither expert was charged with an error. If at least three coders agreed on a code different from the codes of E6 and E7, both were charged with an error. This procedure enabled us to estimate the error rates of the experts. These error rates could then be used to adjust the production error rates as determined by the experts, thus leading to our best measure of the "true" production error rate. We recognize this expert error determination may be somewhat biased in favor of E6 because she had access to the codes of two coders before listing her codes; however, no more suitable measuring procedure was as easily adaptable.

As can be seen in the table which follows, E6 and E7 coded 8,973 records, generating 27,752 code comparisons. These comparisons resulted in the following rates:

(1)	Agreement rate	= 97.76%
(2)	Difference rate	= 2.24%
	a) error rate of E6	= 0.56%
	b) error rate of E7	= 1.26%
	c) unresolved	= 0.49%

The sum of a, b, and c is greater than the difference rate because some differences resulted in errors being charged to both experts.

Agreement Rates Between E6 and E7 and Conversion of Difference Rates to Error Rates, for all Combinations of Single and Multiple Sets Coded by E6 and E7

Coding Set Combination	Number of Records	Number of Codes	Agreement Rate (%)	Difference Rate (%) Charge Error to:			
				E6	E7	E6 and E7	Unresolved
Total	8,973	27,752	97.76	0.47	1.17	0.09	0.49
E6 & E7 one set each	8,289	24,743	98.14	0.39	1.04	0.04	0.38
E6 ≥ 2 sets; E7 = 1 set	280	1,215	95.23	0.91	1.32	0.66	1.89
E6 = 1 set; E7 ≥ 2 sets	290	1,244	93.81	1.45	2.89	0.48	1.37
E6 ≥ 2 sets; E7 ≥ 2 sets	114	550	95.45	0.91	2.91	0.18	0.55

C. Estimate of "True" Error Rate

Although the error rate of E6 is much lower than the error rate of E7, we feel that the estimate of the production error rate as measured by E7 is a closer measure of the "truth." Our rationale is based on the fact that E7 operates independently in arriving at her code selections, while E6, because of her access to the work of two coders, is, minimally at least, subject to their influence. A valid counter argument, of course, is that access to the work of two coders gives the expert a broader perspective on different, acceptable coding strategies, although this argument is not supported by the number of multiple sets coded by the two experts (394 for E6 and 404 for E7). However, with the goal of obtaining an expert truth as completely free of other influences as possible, we decided on E7 as the ultimate determinant.

After the typing and punching errors made by coders 5, E6, and E7 were corrected, the "true" error rate of production coding in the sample was measured at 4.10 percent by E6 and 5.36 percent by E7. These rates, however, include the error rates of the two experts, 0.56 percent and 1.26 percent, respectively. In order to determine what proportion of these error rates should be subtracted from the production error rates in order to get "truth," the errors charged to the experts were reviewed. 8.33 percent of the time that E6 was charged with an error, the production coder agreed with her code. This percentage of E6's error rate was not subtracted from the production coder's error rate. For the same reason, 5.13 percent of E7's error rate was not subtracted from the production coder's error rate. The true production error rate, then, as measured by the two experts, is:

By E6 = 4.10 - (1 - .0833)0.56 = 3.59 percent

By E7 = 5.36 - (1 - .0513)1.26 = 4.16 percent--best measure

D. Comparison of Production Error Rate as Measured by Different Verification Systems

After determining the best measure of the true production error rate in the sample, we were interested in determining which of the following systems,

(1) Three-way independent coding

(2) Two-way independent coding with dependent adjudication of differences

(3) Dependent verification

provides the best estimate of the true error rate.

The three-way independent coding system estimated the production error rate at 3.75 percent. E6 provides the best measure of the error rate that would be obtained under two-way independent coding with dependent adjudication of differences. However, the 4.10 percent referred to above would not be applicable because E6 would review only those code situations in which the production coder and one independent sample coder disagreed. There were 27,952 comparisons between the two independent coders whose work E6 had access to. Of this total, they agreed on 25,897 codes, so there would be no adjudication of these codes. In the difference cases remaining, E6 charged the production coder with 852 errors. Then the production coder's error rate can be estimated by

$$\frac{852}{27,952} = 3.05 \text{ percent.}$$

We encountered a major problem in trying to use the work of coder 5 to estimate the production coder's error rate based on two-way dependent verification. After coder 5's typing and punching errors were removed from the file, she charged the production coder with 1,168 errors out of 27,575 codes, for an error rate of 4.24 percent. This estimate is higher than the estimate based on the three-way system, and is very close to the "true" error rate based on the work of E7. Such a result is, of course, quite surprising because error rates based on dependent verification are expected to be smaller than error rates based on independent verification. In fact, the estimated production error rate for 1973 Mortality Medical Coding, based on a dependent verification system, was 0.7 percent less than the production error rate for 1974 data, measured by independent verification. This deterioration in the error rate occurred despite the fact that the coding staff and the coding instructions were virtually unchanged from 1973 to 1974.

There are two reasons that might explain this higher-than-expected error rate. The first is purely speculative: It is possible that the coders who worked as coder 5, knowing they were working on a special project, were more careful than they would have been if they had been working on a regular data file. The second reason involves a procedural difference between dependent verification of the 1973 file and dependent verification during the experiment. During the experiment, the dependent verifiers were forced to enter one set of codes to represent each certificate. When coding the regular file, dependent verifiers could change the codes of the production coder, yet not charge her with any errors (a process similar to coding two sets of codes, and assigning a "P" to one set). This difference unquestionably caused the production coders in the experiment to be assigned more errors than they would have been if the previous dependent verification procedure had been followed exactly.

In order to estimate the extent of errors charged to the production coder by coder 5 because coder 5 was not permitted to code multiple sets, we examined the multiple sets produced by E6 and E7. The coding of multiple sets by E6 and E7 resulted from ambiguity and confusion as to the correct code for certain medical conditions described and transcribed on the death certificate. It seems reasonable that coder 5, by design less knowledgeable than E6 and E7, would have at least as many ambiguous coding situations as an expert coder. Under this premise, then, coder 5, had she been permitted, would have entered multiple sets for approximately 400 records, 60 percent of which would have had a preferred set, i.e., been given a "P" at the end of the first set. (E6 and E7 designated a preferred set for 59.1 percent and 59.4 percent of their multiple sets, respectively.) These sets represent instances where coder 5 could have overruled the production coder, yet would not have charged her with an error for an acceptable code that she did not feel was the preferred code. Since the procedure followed by E6 is a closer approximation of the one followed by coder 5, the multiple sets of E6 were analyzed. It was determined that 156 codes by the production coder did not match the codes in E6's preferred sets, but were not counted as errors because they were matched in secondary sets. A comparable reduction in the number of errors charged to the production coder by coder 5 yields what we consider to be the best estimate of the expected production error rate based on two-way dependent verification, i.e., $\frac{1168 - 156}{27,575} = 3.67$ percent.

For the thirty work lots included in the experiment, then, we determined that the "true" production error rate is 4.16 percent. The estimate of this "true" error rate from three different coding-verification systems is as follows:

(1) Three-way independent coding--3.75 percent

(2) Two-way independent coding with dependent adjudication of differences--3.05 percent

(3) Dependent verification--3.67 percent

As the above figures indicate, the error rates based on the three-way independent system and the dependent verification system are very similar. This finding contradicts the results of other studies that compared independent verification with dependent verification. While acknowledging the apparent comparability between the two systems under experimental conditions, it should be reemphasized that the awareness of this special study probably influenced the work of coder 5, the dependent coder. Ordinarily the work of a dependent verifier is minimally reviewed. Knowing that all of her work was going to be analyzed may have led coder 5 to perform more diligently than she would have in a normal coding situation. This possible deviation from the normal or expected coding pattern is given additional credence when we consider the work environment of the NCHS coding units. A production standard system exists that places a premium on productivity. That is, the more work that is produced, the more cash remuneration the coder can qualify for. It seems reasonable that a dependent verifier, competing against the clock, would be more inclined to agree with most of the codes listed by the production coder rather than repeat the coding process to determine if she agrees with the listed codes. This, of course, would lead to an underestimate of the production error rate.

Based on these preliminary findings, it appears that dependent verification could estimate the error in Mortality Medical Coding as accurately as three-way independent verification if

(1) The dependent verifier's work was systematically reviewed, and

(2) The current production standards system was revised to place more emphasis on the quality of coding.

Otherwise, the three-way independent system probably provides the best estimate of the quality of the coding.

References

1. Gilford, L., Estimating the Error of the Verifying Clerk when His Work is Not Independent of the Original Clerk's. Sept. 1956. Unpublished paper.
2. McKean, A., Estimating the "True" Process Average by Means of Controller Errors. Unpublished paper.
3. Minton, G., Inspection and Correction Error in Data Processing, Journal of the American Statistical Association, Vol. 64, pp. 1256-1275. Dec. 1969.

SOCIAL STATISTICS AND THE LAW

Philip S. Hart, University of Massachusetts Boston

INTRODUCTION

Statistical decision theory and probability theory stand on the verge of being capable of making significant contributions to legal decision-making. Increasingly more we are seeing statisticians providing the conceptual framework, research designs and expert testimony for cases related to both criminal and civil law. For the most part, those in the legal profession can be regarded as laymen in terms of social measurement. Thus the social statistician must be concerned with presenting findings in a manner and mode which is understandable to an audience of legal personnel. The lesser the ability to communicate findings (and how such findings were derived) then the less likely is it that social statistics will influence legal decision-making. Thus the findings should be both detailed and comprehensible. The graphics accompanying the findings should allow for emphasis and analysis.

In February 1976 issue of the American Statistician, Joseph Van Matre and william Clark note that:

> [However], statistics has not been utilized to any significant degree in some areas; the legal system, particularly the litigation process, would be included in such category...It is nevertheless, the authors' view that the statistician in the role of the expert witness will attract increased utilization from the legal profession in the future. After all, in many respects the trial is simply a search for probabilities. (p.2.)

Van Matre and Clark further note that jury discrimination cases are examples of legal problems that lend themselves to statistical inquiry. Finkelstein goes on to note that "in the more difficult jury discrimintation cases an informed judicial decision cannot be made without mathematical analysis of the underlying data."[1]

In a paper delivered to the 70th Annual Meeting of the American Sociological Association, Nijole V. Benokraitis (1975) reminds us that although social science research has been limited in the area of jury service, there has been some empirical evidence of minimal black participation on juries (Stephenson, 1910; Hearings, 1966; Boags 1971; Broeder, 1965). Benokraitis also notes that since 1966 there has been a tremendous upsurge in filing affirmative jury suits, especially in the South and primarily by the American Civil Liberties Union. Thus, black participation in the jury selection process remains a problematic and controversial issue. This area of concern has been a source of interest primarily for attorneys and jurists, attracting minimal interest from social scientists.

According to David Kairys (1972) successful litigation challenging juror selection procedures arose after the Civil War and its progeny, the fourteenth amendment. Shortly after amendment ratification, the absence of blacks from state jury boxes was challenged on constitutional grounds in cases arising in the border states as well as states making up the Confederacy. Jury discrimination challenges until 1946 dealt with instances of racial exclusion or underrepresentation. The first case to expand the concept of unconstitutional exclusion to include other groups or classes within the population was Thiel v. Southern Pacific Co.[2] Kairys further reminds his readers that:

> Discrimination on the following bases other than race and ethnic background has been prohibited: economic, occupational or social status; religious belief; sex; political beliefs or values; age; and geography (p.780).

Thus, we can see that social statisticians and social scientists have played a limited, yet recently emerging role, in relation to legal decision-making. However, as noted by Van Matre and Clark:

> One may observe that certain types of cases often involve the consultation of statisticians; these include jury discrimination, anti-trust, trademark infringement, and litigation involving injured tort victims (p.2).

The principal focus in this paper will be on jury discrimination cases.

A FOCUS ON JURY DISCRIMINATION CASES

Although everyone generally agrees that a fair jury, fairly chosen, is fundamental to our historic tradtions of justice, a former President of the American Bar Association remarked a few years ago that the subject of jury selection had, "in some inexplicable fashion," escaped the attention of the legal profession.[3] Even judges are often not fully aware of the selection methods employed because they delegated broad powers to the court clerk or jury commissioner; and in practice these officials often operate independently of the judges so as not to impose additional burdens on already overextended courts. As further noted in the Committee on the Operation of the Jury System report:

> The principle that the courts should be vitally concerned with ensuring fair jury selection cannot be challenged. We note in this connection that the President's 1967 Civil Rights message to Congress stated, "creating respect for legal institutions becomes virtually impossible when parts of our judicial system operate unlawfully, or give the appearance of unfairness." (p.17).

In this regard, Kairys notes that the "history of jury selection in this country quite clearly reveals that vast segments of our population have been denied the right to serve on juries. Juries have become representative of the white, middle aged, suburban/rural middle class. Black, poor, and young people, and anyone who sees a need for basic change in the society find virtually no peers on our juries."[4] This phenomenon is further compounded by the fact that the courts have not developed or formalized any measures of discrimination within the legal system. Instead, decisions about discrimination have been based largely on inferences drawn from the results of the jury selection process. Benokraitis notes that:

> In Blackwell v. Thomas (4th Circ., 1973), however, the court accepted a more specific measure of discrimination based on the discrepancy between percentage blacks on the list, panel or box and percent black population: "...a disparity of 10 percent underrepresentation is sufficiently great to warrant an evidentiary exploration of how the jury selection statues are administered." (p.4).

Social scientists are thus playing a larger role in relation to challenging current jury selection procedures at various court levels. This level of applied research is principally concerned with determining the representation of particular cognizable groups in a specified geographic area and comparing this figure with the representation of these same cognizable groups in the juror pool for the same area. The intent then, is to note disparities if they exist, compute the statistical significance of such disparities, and to challenge the selection and representation of the system where such disparities do signal a statistically significant finding.

The focus in jury challenge work is on the source of the jurors, the process of selection, and the results. In isolation, and in toto, these three factors can be used to develop a prima facie case of discrimination in the selection procedure.

The primary audience of work done in jury selection and representation cases are legal personnel including judges, lawyers, court clerks, etc. For the primary audience a particular problem is related to comprehending the intrusion of statistical decision theory into legal decisions. Currently the court system has rather ambiguous standards for determining if a disparity is significant. The social scientist has the conceptual and methodological tools which can contribute to the development of reliable and valid standards. However, as noted by this author in another document[5]:

> A particular problem for the social scientist is related to communicating in an understandable way with the primary audience of legal personnel (p.5).

ROLE OF STATISTICIAN

Relative to social scientists and statisticians working in concert with legal personnel to select "sympathetic" juries a number of charges have been made that the jurors thus chosen are "sociologically loaded dice" and that the procedure amounts to social science jury stacking. One of the initial efforts to make systematic use of the social sciences in jury selection was in federal court at Harrisburg, Pennsylvania, in 1971 and 1972, in conjunction with the trial of Daniel and Philip Berrigan. The social scientists involved designed a four-stage project consisting of the following:

(1) A random telephone survey of 840 residents to determine if the current pool of prospective jurors for the trial actually represented a cross-section of the community.
(2) In-depth interviews with 252 people from the group of 840 to determine the attitudes and characteristics of the types of people likely to show up in the jury pool.
(3) Observing the jurors during the trial.
(4) A follow-up study to be conducted after the jury disbanded to reconstruct how each juror had felt about the defendants and how he had voted.

The other role that social scientists have played is in relation to challenging current jury selection procedures at various court levels. The work of such individuals as: David Kairys in Philadelphia County, Pennsylvania; Hayward Alker in the district court of Eastern Massachusetts; Philip Hart in Suffolk County, Massachusetts; Jay Schulman in Erie County, New York; and George Bardwell in the district court of Eastern Colorado, is instructive in this regard. The work of social scientists and statisticians in jury challenge cases has principally consisted of developing the conceptual framework, research designs, providing expert testimony, and supervising graphics presentations.

Van Matre and Clark further note that in addition to actually testifying as an expert witness the statistician may also be called upon to provide other expert assistance

> ...which is equally as valuable, such as (1) listen to the cross-examination; (2) study the deposition of an opposing expert; (3) assist the examiner in preparing questions for the opposing expert; (4) prepare reports explaining sampling and other statistical issues (5) collect and analyze related data derived from an independent source; (6) give the examiner positive and clear-cut recommendations and decisions with regard to such basic aspects as the adequacy of the sampling plan and the trustworthiness of the data derived from the sample (p.4).

As a specific example of tasks associated with jury challenge work, the following outlines the study procedures in Suffolk County, Massachusetts.[6] The first step in the study was to establish a research design that would allow systematic study of the actual results of the juror selection process. Specifically, we wished to compare each of the relevant populations, i.e., census, resident list, jury lists, and persons actually summoned for jury duty for the most recent possible year.

The next step was to translate the 1970 census tract data into ward data. While the census tracts often coincide with wards, often they do not. To facilitate comparison with the resident list which is compiled by wards, the study team recompiled the census data by ward. Next a one percent representative sample of the 1973 resident list was drawn. The sample was drawn pursuant to the order of Judge Kent Smith which granted the study team access to the city computer resources. Next a 5 percent random sample of the 1973 jury list was drawn. The final sampling step was to draw a 5 percent sample of the September 1973 to June 1974 juror pool. The sampling results were then compared with cognizable group representation in the universe comprised of the 1970 U.S. Census and statistical tests applied at the .01 level to determine whether disparities found were consistent with chance occurrence. Social graphics were then prepared and a courtroom presentation made which in combination with the Supreme Court ruling on barring women from jury duty, facilitated changes in the Suffolk County juror selection process.

Once the statistician-as-expert has satisfied himself that the area of interest is one which calls for his particular expertise, he must then prepare for his testimony. Careful cooperation between the expert and lawyer is essential to efficient use of the statistician-as-expert-witness. The questions directed to the expert in the courtroom will either be hypothetical or based upon the actual facts. In sampling-related testimony, the expert might be asked such questions as:

1. How can you tell when you have good results and when you have bad results from a sample?
2. Did your client give you any instructions regarding what the results of your study were expected to be?
3. How was the sample size ascertained?
4. A wide confidence interval is really a large margin of error, isn't it?

David Kairys notes that the role of the expert is not confined to checking the accuracy of the mathematical computations. More importantly, an expert is necessary to explain the principles involved and to present the statistical evidence. The presiding judge must be convinced that the statistical principles are valid and be persuaded to receive the evidence. A court may balk at expert testimony which appears to be conclusive of the legal issues.

The expert witness must also be concerned with and aware of the ethical implications attendant to such work. The statistician faces the dilemma of whether he has an ethical responsibility to be totally neutral. Gibbons points out that:

> It is essential that the statistician inform his employer of his neutral position on all strictly non-statistical aspects of the study before agreeing to undertake an investigation, as his position as an independent agent is considerably weaker once the study commences. (p.74).

The question of fees is also related to questions of ethical considerations. The statistician may consider serving as an expert witness on a contingent fee basis. However, one's objectivity may be questioned by the opposing attorney and jury if the expert has a financial stake in the outcome. The statistician should not act as an advocate, that is the lawyer's role. The statistician-as-expert witness can best serve himself and his employer by being neutral.

PUTTING TOGETHER THE REPORT

As noted earlier, to date the courts have not explicity defined at what point a disparity is significant. The role of the statistician should be to aid the courts in establishing standards which have as their basis sampling theory and statistical decision theory.

Three indispensable variables must be accounted for in determining the significance of a disparity. These variables are:

(1) The quantitative disparity between a group's representation in the population and the jury pool.
(2) The size of the sample used to determine the proportions of the pool.
(3) The range at which the disparity occurs.

Further, as Kairys notes in this regard:

> There are simple, mathematically precise methods for making this kind of comparision and accounting for all relevant variables. These methods are neither new to the science nor so complex or advanced as to be impractical for use by the courts...
>
> The mathematical method provides a means of defining the point at which a disparity between the population and the juror pool becomes significant by a computation of that disparity resulting if the process were in fact unbiased. (p.786).

The statistical inquiry thus usually involves sampling techniques, inferential statistics including non-parametrics, and partial correlation analysis. The chi square and binomial distribution are computations commonly used in jury challenge work.

However, the statistician must be concerned with these simple, yet precise, methods playing a role in legal decision-making. Mathematician, film-maker and graphic artist Ugo Torricelli has outlined a seven-step approach to visualizing complex statistical data[8] which is of import here. The approach is outlined as follows:

(1) Interaction
(2) Graphic exposition
(3) A two-dimensional visual field
(4) A three-dimensional visualization
(5) Dynamic evolution
(6) Collateral documentation
(7) Dramatization

The goal of this visualization of the simple, yet precise, statistical data generated in jury challenge work would be to enhance comprehension on the part of legal personnel as laymen in social measurement. For as Katz states in the February 1975 issue of the American Statistician:[9]

> It is to be hoped that in the future attorneys and judges will become more knowledgeable about chance, uncertainty, probability, statistical procedures,

> and statistical inference in the presence of uncertainty, so that the instructional phase of the statistician's testimony might be shortened (p.142).

Thus the report, and testimony, should be addressed to making clear the findings, how the findings were derived, and the accompanying tables, charts, maps, and graphs. The oral and written word should emphasize visualization and comprehension.

For example, in terms of sample selection the goal should be to explain the theory and process of sampling, in addition to attending to questions of the sample's relevance to the universe from which it was drawn. Katz notes that the presentation of the estimates of the prevalence of some condition in a large population derived from a sample survey requires extremely stringent application of the principles of sample selection, particularly of the treatment of non-response.

> Traditionally, opposition counsel have been quick to attack open possibilities, regardless of likelihood, as conceivable alternatives (Katz, p.138).

The social scientist or statistician should then be prepared with visual aids depicting the sampling process and sampling theory. Such visual aids as part of a courtroom demonstration should be reviewed with the attorney prior to courtroom implementation. This method of continuous interaction with legal personnel should begin with the employing attorney and continue through interaction with the presiding judge and cross examining attorney. The goal of the statistician in a neutral role should be to enhance comprehension on the part of all legal personnel concerned.

Likewise, an important consideration in any case is whether evidence derived from a sample survey is admissable under existing rules of evidence, the principal objection having been that the evidence is hearsay. As Katz notes, courts now admit samples or polls over the hearsay objection (a) on grounds that surveys are not hearsay and (b) on the grounds that surveys are within a recognized exception to the hearsay rule, or (c) without stating the grounds for admitting a survey.

The rules of evidence are thus important to know and understand by the statistician. Whether evidence is admissable if gathered through a sample survey, telephone polling, estimation procedure, etc., are items to know prior to beginning work. In this regard, the statistician should be aware of the distinction made between a sample of "objectively observable facts" and a poll of "views or attitudes."

The report should also clearly explain statistical tests of significance. A tendency here on the part of the statistician to utilize more elegant tests may work to his disadvantage. This problem may be addressed by finding the manner and mode of combining scholarly accuracy and complexity with the need of communicating results to audiences of non-specialists in social measurement.[10] A more dynamic visualization of statistical tests and significance levels may provide a bridge between these antithetical stools.

Well conceived and designed social graphics can be useful in terms of aiding the visualization process. Once the source, process, and results have been determined, these findings can be formulated into a social graphic presentation aimed at making the selection process and resultant data comprehensible to the court. As noted by the author in a recent monograph:

> For the most part the graphics presented in jury challenge cases to date have been fairly standard two-dimensional bar graphs, charts and tables. Because of the increasing importance in encouraging the court system to set standards for disparity, social scientists should become more aware of various forms and mediums for graphic presentation. An assumption here is that with more visual graphics comes a heightened level of comprehension, thus increasing the likelihood that standards for disparity will be adopted in the court system. (p.7).

CONCLUSION

The presentation of statistical evidence to a court composed of statistical laymen is often a challenge in itself. It is hoped that in the future attorneys and judges will become more knowledgeable about statistical decision theory and probability theory. Social graphics which enhance visualization may contribute to bringing about this knowledge. The findings should be both detailed and comprehensible. The graphics accompanying the findings should allow for emphasis and analysis.

In that legal personnel have primarily been in the role of information receivers, the social graphic format and medium should be designed so as to expand this role. For it will be at the point of comprehending the data and manipulating the data that legal personnel can "supply" the standards necessary to move the criteria for decision-making from the subjective and intuitive to the objective and quantifiable. As Benokraitis notes:

> Since the "causes" of racially unrepresentative juries are seen to be residing in institutions rather than individuals, remedies for change must also be focused on institutional and situational arrangements, rather than seeking corrective programs or strategies for individual characteristics (p.18).

Thus the applied use of social statistics discussed in this paper can move to facilitate institutional change in the court systems. A prudent, clear, and effective use of social graphics can contribute to this movement to adopt objective standards for jury selection procedures.

NOTES

1. Finkelstein, Michael O. "The Application of

Statistical Decision Theory to the Jury Discrimination Cases," Harvard Law Review, Vol. 80, 1966, p.340.

2. 328 U.S. 217 (1946).
3. As noted in The Jury System in the Federal Courts. Works of the Committee on the operation of the Jury System of the Judicial Conference of the United States (1966-1973), p.17. 1973.
4. Kairys, David. "Juror Selection: The Law, A Mathematical Method of Analysis, and a Case Study," The American Criminal Law Review. Vol. 10, no. 4. (Reprint). p.801. Dec. 1972.
5. Hart, Philip S. Toward Social Graphic Approaches for Jury Selection and Representation Problems. Monograph prepared with support of the Bureau of Social Science Research, Inc., Graphic Social Reporting Project. May 1975, p.5.
6. Report of Jury Study Suffolk County. Commonwealth of Massachusetts, Superior Court Nos. 80793, 80794, and No. 75662, Dec. 1974.
7. Sources: Joseph G. Van Matre and Wm. N. Clark, "The Statistician as Expert Witness," The American Statistician, v. 30, no. 1, Feb. 1976; and transcript from the Hearing on Amended Motion to Dismiss the Indictment and to Strike Traverse Jury Venire. Commonwealth of Massachusetts, Suffolk Superior Court, December 2, 1974.
8. As reported in Feinberg, Barry M., Kinostatistics: Communicating a Social Report to the Nation. Bureau of Social Science Research, Inc. December 1972.
9. Katz, Leo. "Presentation of a Confidence Interval Estimate as Evidence in a Legal Proceeding," The American Statistician, v. 29, no. 4, November 1975, pp. 138-142.
10. See Biderman, Albert D. "Kinostatistics for Social Indicators," Educational Broadcasting Review, v. 5, no. 4. (Reprinted from October, 1971.) Biderman notes that an "important source of the failure (of social reporting) appears to be the tendency of the attempts to fall between two antithetical stools..." (p.14).

REFERENCES

1. Alker, H.R. Jr. et al The Jury Selection Process in Eastern Massachusetts. (Preliminary Version). Dept. of Political Science, Massachusetts Institute of Technology, 1974.
2. Benokraitis, N.V. Exclusion of Blacks from Jury Selection Process in the South: Some Preliminary Analyses. Paper presented to 70th Annual Meeting of the American Sociological Association, August 1975.
3. Biderman, A.D. Kinostatistics for Social Indicators, Educational Broadcasting Review, v.5, no. 4. Reprinted from October 1971.
4. Brown, K.S. and Kelly, D.G. Playing the Percentages and The Law of Evidence, Law Forum, Vol. 1970, No. 1, pp. 23-48.
5. Burnham, Stern and Shapiro, Attys., Boston Report of Jury Study, Suffolk County, December 1974. Boston, Massachusetts.
6. Committee on the Operation of the Jury System of the Judicial Conference of the United States. The Jury System in the Federal Courts (1966-1973). Compliments of West Publishing Company. 1973.
7. Commonwealth of Massachusetts Hearing on Amended Motion to Dismiss the Indictment and to Strike Traverse Jury Venire. December 2, 1974. Boston, Massachusetts.
8. Cowan, T.A. Decision Theory in Law, Science, and Technology, Rutgers Law Review, v. 17., 1963.
9. Des Raj. Sampling Theory. McGraw-Hill, New York, 1968.
10. Erlanger, H.S. Jury Research in America, Law and Society Review. 4 (February): 345-370.
11. Feinberg, B.M. Kinostatistics: Communicating a Social Report to the Nation. Prepared under NSF Grant GS-29115. Bureau of Social Science Research, Inc. December 1972.
12. Finkelstein, M.O. The Application of Statistical Decision Theory to the Jury Discrimination Cases, Harvard Law Review, vol. 80, 1966, pp. 338-376.
13. Gibbons, J.D. A Question of Ethics, The American Statistician, v. 27, 1973, pp. 72-76.
14. Hart, P.S. Toward Social Graphic Approaches for Jury Selection and Representation Problems. Prepared in conjunction with Bureau of Social Science Research, Inc. May 1975.
15. Kairys, D. Juror Selection: The Law, A Mathematical Method of Analysis, and A Case Study, The American Criminal Law Review. Reprinted v. 10, no. 4, 1972.
16. Katz, L. Presentation of a Confidence Interval Estimate as Evidence in a Legal Proceeding, The American Statistician, v. 29, no. 4, Nov. 1975. pp. 138-142.
17. ------------ Confidence Intervals for the Number Showing a Certain Characteristic in in a Population When Sampling Without Replacement, Journal of the American Statistical Association, 48 (1953), 256-261.
18. Rhine, J. The Jury: A Reflection of the Prejudices of the Community, Hastings Law Journal, v. 20, 1969, pp. 1417-1445.
19. Robinson, W.S. Bias, Probability and Trial by Jury, American Sociological Review, v. 15, 1950, pp. 73-78.
20. Talbot, T.F. Your Day in Court, Machine Design, Feb. 6. 1975, pp. 68-72.
21. Van Matre, J.G. and Clark, W.N. The Statistician As Expert Witness, The American Statistician, v. 30, no. 1, Feb. 1976. 2-5.
22. Wigmore, J.G. Wigmore on Evidence. Boston: Little Brown and Co. 3rd Ed. 1940:

Note: This paper was accompanied by a 5-minute slide show depicting social graphics presented by social statisticians in various court cases.

The U - Hypothesis Revisited

Robert M. Hashway, Boston College

Abstract

In their 1975 report Nuttall and Nuttall (1975) discuss the results of an intensive study of family characteristics in Puerto Rico, and Boston. The method of gathering data that they used was for the most part personal interviews with the respondents. Nuttall and Nuttall (1975) report a 'U-shaped' profile between family size and socioeconomic status. The research reported in this paper was an attempt to determine if this profile would emerge with a diferent sample, and a different data acqisition technique. The sample consisted of students from randomly selected classrooms in nine schools selected on the basis of the average socioeconomic status of the community, and community size. The communities sampled were from Rhode Island, and eastern Massachusetts. The students responded to a questionnaire with the size of their family (number of children including themselves) and the occupation and education levels of their parents. The socioeconomic status variable was calculated by adding the z-scores of the parent's education and occupations, and dividing this sum by four. The SES score of a family served as the basis for the definition of eight SES catagories. It was found that the U-shaped profile found in this data possessed a highly significant linear component. The departures from linearity were due to lower middle, and lower lower class families. The lower middle class families behaved in s fashion similar to upper lower class families, and lower class families behaved as lower central middle class families with respect to family size.

INTRODUCTION: Nuttall and Nuttall (1975) report the results of an intensive investigation of the characteristics of Puerto Rican, and suburban Boston families. One of the many results reported in that study was a 'U-shaped' relation between family socioeconomic status and family size. These investigators report that: "very low and very high socioeconomic status families will have a tendancy to be large, while middle status families will tend to be small according to the U-shaped hypothesis" (Nuttall and Nuttall 1975, p. 68). The purpose of this research was to determine if the U-shaped SES-family size profile would emerge from data obtained from a different sample, and method of data collection.

METHOD: The data used in the Nuttall and Nuttall (1975) study was gathered through personal interviews. The data for this study was collected using a survey instrument. The sample consisted of the families of 909 children selected from nine school districts in Rhode Island and Massachusetts. The school districts were choosen as so that upper, middle, and lower SES as well as large and small communities were equally represented. Within each school classrooms were randomly selected. Only one grade level within each high school was selected inorder to reduce the probability that large families would be over-represented. Each student in each of the classrooms completed a short questionnaire. These students were asked to list the occupation, and highest level of education achieved by each of their parents, and the number of children in the subject's family (including the subject him or her self).

The education level of each of the parents were catagorized according to the scheme described in Table 1. The occupation levels of each of the parents were coded using a modified Warner scale (Inkeles & Smith, 1974). Occupations were coded on an eight point scale. The higher the number the greater the occupational status. A code of 8 would be assigned to college professors, bank executives, owners of large businesses, doctors, lawyers, etc.. A score of 1 would be assigned to unemployed individuals, or wives (or husbands) who were 'at home'.

A global socioeconomic status measure was calculated by adding the z-score of the occupation, and education levels of the parents, and dividing this sum by 4. This procedure should, if the variables were independent, result in a mean score of 0.0, and a standard deviation of 1.0. The observed mean was -.002. Since the parent's occupations and education levels were correlated, the observed standard deviation of .684 was somewhat less than the theoretical expectation assuming independance. The mean education, occupation, and SES levels are reported for each family size grouping in table 2. It was found that the observed variation in each of these indices could not be attributed to family size. Using each of these indices as dependent variables in a oneway analysis of variance failed to result in a significant 'main effect' at the .05 significance level.

The global SES parameter was recoded into eight catagories. If a family had an SES score of x, it would fall into the class $INT(4 + (2x/s_{SES}))$. This scheme made it possible to divide the upper, and lower classes into two groups each, and the middle class into four groups. The terms assigned to these groupings are standard terms most commonly found in the socioeconomic literature (Broom & Selznick, 1963). This procedure is similar to that used by Nuttall and Nuttall (1975).

RESULTS: A oneway analysis of variance procedure with the family size as the dependent variable, and the SES catagorization as the independent variable resulted in a large main effect (F=2.682; df=7,901; p less than .01). The mean family size gor each of the SES groups are shown in table 3, and figure 1.

An inspection of the graph in figure 1 would indicate that a U-shaped curve similar to that found by Nuttall and Nuttall (1975) exists with this data. However, since this population was distributed in such a way as to facilitate a finer division of the middle class, the shape of the profile in figure 1 is somewhat asymmetric. The lower middle class families seem to have the least number of children. The lower middle class families seem to be similar to the upper lower class families in terms of family size. The lower lower class families seem to be similar to the lower central middle class families in terms of the number of children.

The profile shown in figure 1 seems to be parabolic in character. This hypothesis was tested using a polynomial regression procedure (Winer, 1971). The results of this analysis is described in table 4. These results indicate that only the linear term can be considered to be significant (F=10.128; df=1,901; p less than .002). All higher order terms up to the quintic (degree 5) were found not to be significant at the .05 level. The linear term accounted for 54 percent of the variance in the family size (multiple correlation of .736). It can be easily seen from figure 1 that the departure from linearity is primarily due to the influence of the lower middle, and lower class families. Lower class families seem to be more like lower central middle class families than would be expected from a linear model. Lower middle class families seem to behave more like upper lower class families than one would expect from a purely linear model.

DISCUSSION: The main objective of this research was to demonstrate that the U-shaped profile between family size and SES reported by Nuttall and Nuttall (1975) could be found in other samples, and using other techniques for data acquisition. The existance of the U-shaped profile is called the 'U-Hypothesis' by Nuttall and Nuttall (1975). The results of the analysis of the data collected from this sample, using the child self-report questionnaire, indicates that the U-shaped profile can be found from different samples and using other procedures.

The U-shape seems to be more of a departure from linearity than a parabolic tendency in the population. Lower middle class families seem to behave more like upper lower class families, and lower class families seem to behave more like lower central middle class families. The behavior of other social classes with regard to family size can be explained by a linear SES by family size profile. Except for lower middle class, and lower class families, it seems that the higher a families SES the more children it can and does support.

The fact that lower middle class families are similar to upper lower class families is not surprizing. Assuming that upward social mobility is operating here, the lower middle class family can be thought of as being 'derived' from an upper lower class parentage (or grandparentage). Therefore, it would be expected that some of the child bearing characteristics of the upper lower class would be reflected in the lower middle class.

The general linear profile would imply that the higher the SES level of the family the larger the family size. The similarity between lower lower class families and lower central middle class families can be explained in terms of the linear model if we assume that higher SES means a somewhat larger family income. The lower lower class family in this sample were either unemployed, or employed at a very low level. It seems likely that they would be obtaining some type of public assistance. Given the large number of public agencies giving support to the underprivleged, it seems most likely that in terms of finances the lower lower class family could be similar to the lower central middle class family. Therefore, one would expect a similarity to exist between these social classes with regard to the size of their families.

REFERENCES:

Broom, L. and Selznick, P. Sociology, 3 rd. ed. New York: Harper & Row, 1963.

Inkeles, A. & Smith, D. Becoming Modern. Cambridge, Ma.: Harvard University Press, 1974.

Nuttall, R. L. and Nuttall, E. V. FAMILY SIZE AND SPACING IN THE UNITED STATES AND PUERTO RICO. Final Report Contract NICHD-72-2033 11-22-71. Chestnut Hill, Ma.: Laboratory For Statistical And Policy Research, Boston College, March 1975.

Winer, B. J. Statistical Principles In Experimental Design, 2nd. ed.. New York: McGraw-Hill Book Company, 1971.

Table 1

Education Catagorizations

9. Phd., Edd., LLd., M.D., DDS.
8. Master's Degree
7. B.A. or B.S. and more college
6. B.A. or B.S. only
5. Some College
4. High School Graduate
3. Some High School
2. Jr. High School Graduate
1. Elementary School Only

Table 2

Mean Education And Occupation By Family Size

		FAMILY SIZE (NUMBER OF CHILDREN)							
		1	2	3	4	5	6	7	8
Mother's Occupation	Mean	2.919	3.209	2.946	2.75	2.888	2.538	2.533	2.197
	Std. Dev.	2.465	2.530	2.413	2.431	2.383	2.396	2.646	2.294
Education	Mean	2.865	2.550	2.778	2.631	2.559	2.478	2.220	2.051
	Std. Dev.	1.030	1.953	1.962	2.044	1.905	1.915	1.920	2.224
Father's Education	Mean	3.086	3.033	3.168	3.449	2.946	3.067	3.291	2.831
	Std. Dev.	2.418	2.110	2.160	2.064	2.111	2.297	2.043	2.386
Occupation	Mean	4.056	4.000	4.019	4.173	3.656	3.713	3.411	3.902
	Std. Dev.	2.317	2.201	2.344	2.187	2.265	2.363	2.104	2.420
Family S.E.S.	Mean	.042	.030	.070	.062	-.055	-.043	-.098	-.161
	Std. Dev.	.647	.685	.648	.656	.666	.675	.634	.771

Table 3

Mean Family Size By Social Class

Social Class	N	Mean	Std. Dev.
Upper Class:			
Upper Half	66	5.1515	1.8080
Lower Half	76	3.9342	1.9207
Middle Class:			
Upper	114	3.8509	1.6893
Central:			
Upper	92	4.1957	1.7866
Lower	157	4.2611	1.8985
Lower	123	4.3984	1.9066
Lower Class:			
Upper	98	4.4286	1.8667
Lower	183	4.6831	1.9522

Table 4

Polynomial Regression Results

Source	D.F.	SS	MS	F	R^2
Between Groups	7	65.4063	9.3438	2.682*	N.A.
Linear Term	1	35.3934	35.3934	10.128**	.541
Dev. From Linear	6	30.0128	5.0021	1.436	N.A.
Quadratic Term	1	6.2280	6.2280	1.784*	.095
Dev. From Quadratic	5	23.7849	4.7570	1.365	N.A.
Cubic Term	1	2.9982	2.9982	.859*	.046
Dev. From Cubic	4	20.7867	5.1967	1.491	N.A.
Quartic Term	1	3.0973	3.0973	.887*	.047
Dev. From Quartic	3	17.6893	5.8964	1.692	N.A.
Quintic Term	1	.3150	.3150	.090*	.005
Dev. From Quintic	2	17.3743	8.6872	2.493	N.A.

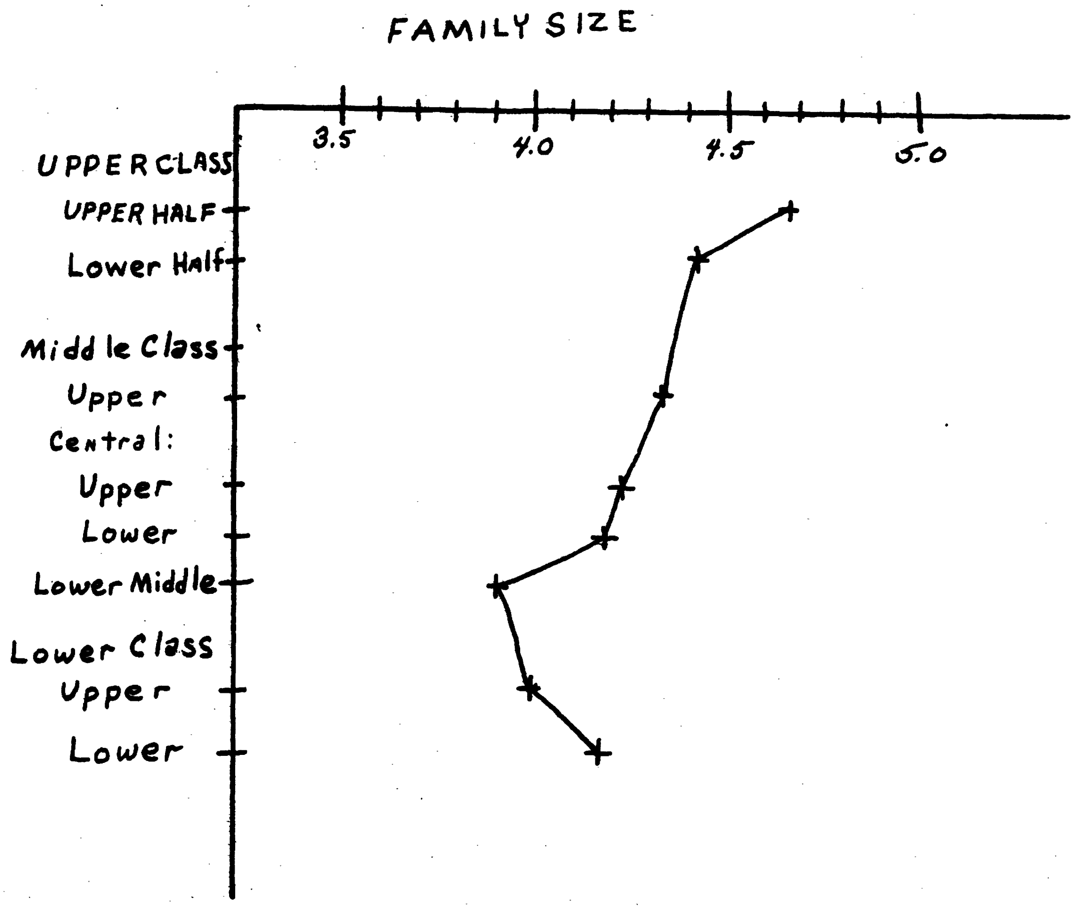
* $p < .001$

* $p > .05$

N.A. Not Applicable

Figure 1

SES By Family Size Profile



Factors Affecting, and Patterns Between Quarterly School Absence Rates

Robert M. Hashway, Boston College

Abstract

Recently, some attention has been directed toward the affects of the quantity of schooling on achievement (Karaweit, N. 1973, 1975; Kidder, O'Reully, and Kiesling, 1975; Wiley 1973, 1974). This paper reports the results of a preliminary investigation of the effects of sex, socioeconomic status, family size, and a student's level of occupational aspiration on within school quarterly absence rates. The sample consisted of 167 ninth grade students randomly selected from the high school population of a suburban northeastern community. It was found that the sex of the child did not significantly affect the number of days the child was absent in any one quarter, or for the entire year. The size or socioeconomic status of the child's family, and the child's level of occupational aspiration had different impacts upon attendance at different times during the school year. Of the variance explained by all of the variables, 35 to 38 percent could be attributed to the characteristics of the home (family size, or socioeconomic status). The best predictor of student absence rates during a particular quarter was the student's previous history of absence. Multiple regression analysis was used to determine the relationships between background and previous attendance history.

INTRODUCTION: Many studies of school or program effects have arrived at somewhat less than positive results. These studies tend to indicate that more of the variance in achievement measures are attributable to family background variables than to between schools or programs. Recently, some attention has been directed toward the affect of school attendance on achievement (Karaweit, N., 1973, 1975; Kidder, O'Reully, and Kiesling, 1975; Wiley, 1973; Wiley and Harnischfeger, 1974; Harnischfeger and Wiley 1975). The rationale seems to be that the longer a child is in contact with the school environment (days of attendance) the more he or she will learn. The purpose of the research reported in this paper is to explore the relationships which may exist between school attendance and the characteristics of the family.

METHOD: The sample consisted of 167 ninth grade students from a suburban northeastern school system. The system used the 6-2-4 grade system. Therefore, these students were all entering high school freshmen. The students were randomly selected from a class of about 400. The sample contained 70 female, and 97 male students. Family background, and attendance data were obtained from existing school records. The occupations, and education levels of the student's parents were coded according to the scheme shown in tables A and B. The child's career aspirations were coded on the same scale as the parent occupations. The size of the student's family was recorded as the number of children (including the subject) currently living in the same abode as the subject.

An index of socioeconomic status was computed. This index was the average of the z-scores of the parent's education, and occupation. In theory, if the variables were independent the averaging process would result in a variable with a mean of zero and a variance of one. The observed mean was -.004. Since the occupation and education levels of the parents were highly correlated (about .5 to .54 with each other) the observed variance of .86 was somewhat lower than the expected value of 1.0.

The number of days which the students were absent from school during each quarter, and the total number of days absent during the school year were used as criterion variables. Multiple regression techniques were used to determine the effects of the home background variables on school absence.

RESULTS: Seperate multiple regression analyses were performed to assess the amount of variance in each quarter's absence rate, and in the total number of days absent that was explained by the home background variables, and previous attendance history. The results of each of these analyses are described in Table 1.

Some interesting observations can be made from the entries in Table 1. First, the student's level of occupational aspiration seemed to have had an effect upon the number of days of schooling the student received. The sign of the beta weight between aspiration and attendance indicated that students with higher levels of aspiration tended to have fewer days absent in the first quarter. However, after the first term the effect of the aspiration variable was not significant. The student's aspiration level did not seem to significantly effect the total number of days absent after the first term.

The size of a student's family seemed to have an affect on school absence accross quarters, except for the last (4 th. quarter) term. Students from larger families tended to be absent from school more often than their peers from smaller families.

The socioeconomic status of the family did not seem to effect the number of days which the student was absent except in the last term. There are numerous hypotheses which may explain this phenomina. In the last term most families, regardless of size, realize that this is the last term in which their children can make up deficiencies which may have resulted from absence or some other source. Therefore, more attention may be paid to end of term three reports than to reports from terms one, or two. The importance of the last term may serve as an incentive for parents of large families to pay more attention to the individual child's actions than they may have, otherwise. Therefore, the effect of family size would be diminished. Since the family size variable was highly correlated with the SES variable (r=.659), the reduction of the influence of family size would release more of the criterion's variance for the SES variable to explain. The data necessary to investigate this hypothesis, or others was not present. The inference, however, seems to be that children from high SES families tend to be absent less in the last term than children from low SES families. This can be explained by numerous causal hypotheses. The least important of which is not the social component of schooling. It is well known that a large amount of the school's social activity occurs in the last quarter of the year. It seems reasonable to assume that children from high SES families would be concerned with these activities. To be absent from school would imply a partial reduction in the number of social activities which a student would be actively involved. Therefore, higher SES students would be expected to be in school more often, and absent less than lower SES students. This is not the only hypothesis that would explain this effect. Many other theories can be developed which would tend to explain the observed phenominon. It is probable that some combination of these theories is at work.

The results presented in table 1 with regard to the total number of days absent yields a very clear picture. The major and only significant family background factor whcih seems to affect the absence rate is the size of the family with whom the student resides. Students from larger families can be expected to be absent more during the school year than students from smaller families.

Indications of other interesting phenomina can be seen from the results reported in Table 1. The largest significant predictor of a student's absence rate in a particular term seems to be the child's past history of absence. Also, the predictive effect of term one's attendance seems to diminish after term two. The number of days absent in term four seems to be dependent upon the number of days a student was absent in terms two, and three but not in term one. The directionality of the effects are shown in figure 1. The directionality is shown by the arrows, and the numbers are the beta weights of all predictors with a significance level below .05.

The fact that term one does not effect terms three, or four is not surprizing. In term one, and the beginning of term two school is still a 'new' or novel experience after the long summer. Students are meeting old friends, and making new ones. New and interesting social relationships are being formed. However, as the school year continues, the novel experiences of term one become the day to day routine of the school climate.

A final result which is interesting since it is somewhat counter intuitive is the insignificance of the sex variable. One might expect that the 'Tom Sawyer' effect should be present. Intuitively, it would seem that boys might be more interested in sports, fishing, hunting, etc., than in school events. This theory, if correct, would imply that boys would exhibit higher absence rates than girls. However, it seems that boys and girls do not differ significantly with regard to the number of days which they attend school.

This sample consisted of students from one grade in a particular school system. The results can not be generalized to the population of all schools. However, the system sampled for this study was not much different from other suburban school systems in the industrial Northeast. The results obtained from this study should be similar to what one might expect to find in other comparable towns. Some interesting observations were: the effects of past attendance history, family size and SES on quarterly attendance, and than 35 - 38 percent of the total amount of explained variance was attributed to family background factors. The results of this research served to indicate what might be expected to be found in other suburban, and possibly urban towns.

REFERENCES

Karweit, Nancy. Rainy Days and Mondays: An Analysis of Factors Related to Absence from School. Baltimore: Center For Social Organization of Schools. The John Hopkins University. Report No. 162, November, 1973.

______________. Quantity of Schooling: A Major Educational Factor? Educational Researcher, Vol. 5, No. 2, Feb. 1976.

Kidder, S. J., O'Reully, R. P. and Kiesling, H.J. Quantity and Quality of Instruction: Empirical Investigations. Paper presented at the A.E.R.A. meeting, Washington, D. C., March 31 - April 3, 1975.

Wiley, D. E. Another Hour, Another Day: Quantity of Schooling, A Potent Path for Policy. University of Chicago. Studies of Educative Processes. Report No. 3, July 1973.

Wiley, D.E. and Harnischfeger, Annegret. Explosion of a Myth: Quality of Schooling and Exposure to Instruction, Major Educational Vehicles. Educ. Researcher, April, 1974.

Harnischfeger, Annegret and Wiley, D.E. Schooling Cutbacks And Achievement Declines: Can We Afford Them?, The University of Chicago, Administrator's Notebook, No. 1, 1975.

Table 1

Multiple Regression Results

	BETA WEIGHTS				
INDEPENDENT VARIABLES	TERM 1	TERM 2	TERM 3	TERM 4	TOTAL
SEX	-.04	-.00	-.07	-.05	-.01
SOCIOECONOMIC STATUS	-.04	-.08	-.01	-.19*	-.02
FAMILY SIZE	.22*	.26**	.18*	-.00	.39**
Asperation Level	-.24**	-.02	-.02	-.07	-.16
Absence Term 1	N.A.	.54**	.04	-.04	N.A.
Absence Term 2	N.A.	N.A.	.61**	.28**	N.A.
Absence Term 3	N.A.	N.A.	N.A.	.31**	N.A.
Total Variance Accounted For:	12.7%	41.0%	54.6%	33.0%	21.3%
Percent Of Total Accounted Variance Explained By Family Background Variables	N.A.	38%	38%	35%	N.A.

N.A. Not Applicable

* $p < .05$

** $p < .01$

FIGURE 1
RELATIONS BETWEEN VARIABLES

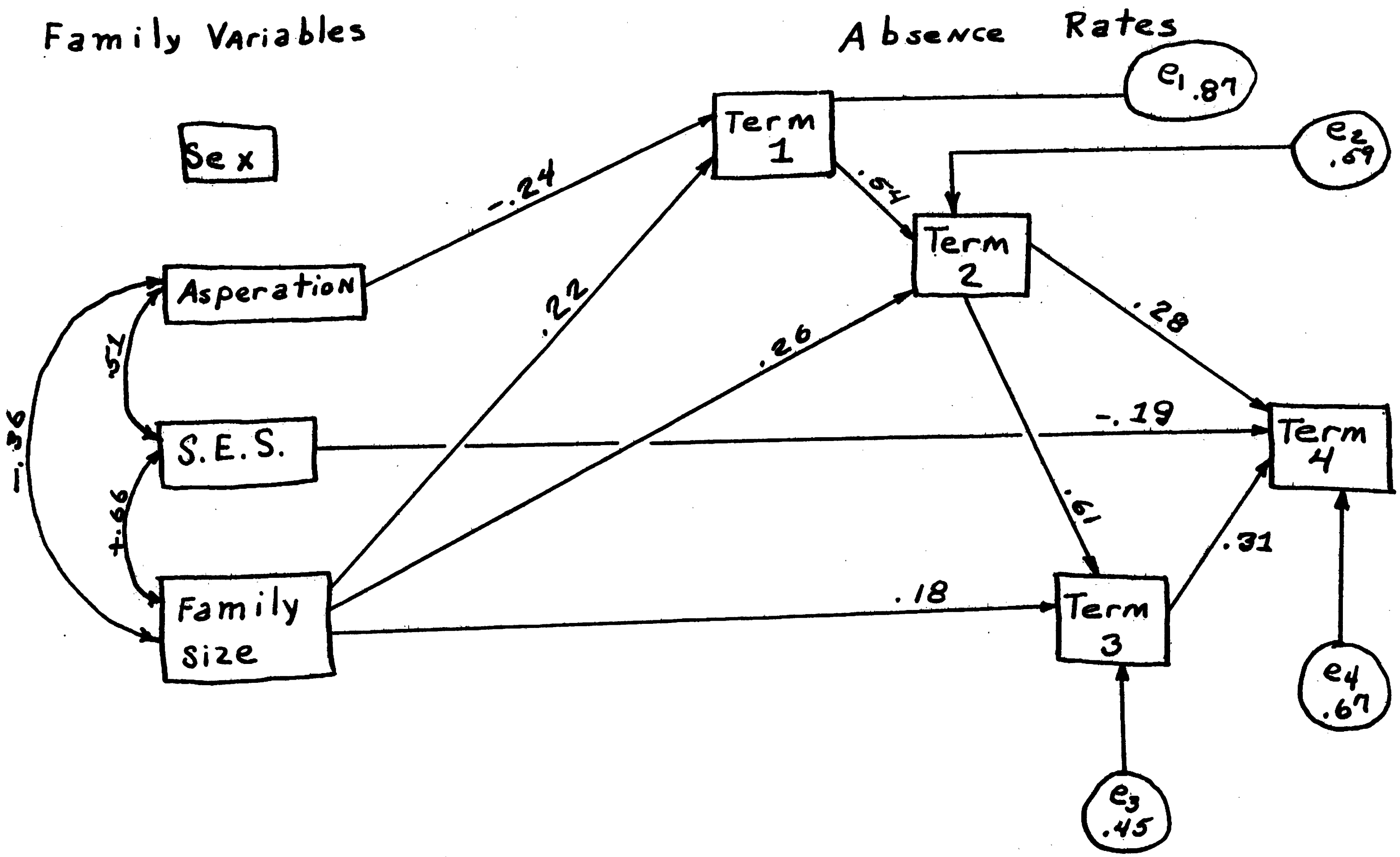


Table A

Occupational Classifications

Group 8:	Architect, Banker, College Professor, Doctor, Engineer, Judge, Lawyer, Manager of a Business, Officer in the Armed Services, Owner of a Large Business, President of a Large Corporation, Town or State Offical.
Group 7:	Artist, Computer Programmer, Entertainer, High School or Jr. High School Teacher, Minister, Newspaper Reporter, Nurse, Pharmacist, Real Estate or Insurance Salesman, Social Worker, Surveyor, Undertaker, Veternarian, Writer,
Group 6:	College Student, Contractor, Construction Supervisor, Draftsman, Foreman, Fireman, Owner of a Small Business, Policeman, Professional Athelete, Radio/T.V. repairman, Salesman, Secretary, Supervisor, Technician.
Group 5:	Electrician, Elementary School Teacher, Machinist, Office Clerk, Stenographer, Tool Maker.
Group 4:	Baker, Carpenter, Cobler, Mechanic, Plumber, Printer, Steel Construction Worker.
Group 3:	Corporal or Private in Armed Services, Bartender, Clerk, Gas Station Attendent, General Factory Worker, Hoist Operator, Machine Operator, Waitress.
Group 2:	Barber, Brick Layer, Bus Driver, Butcher, Butler, Cook, Housewife, Janitor, Laborer, Maid, Night Watchmen, Taxi Driver, Truck Driver.
Group 1:	Unemployed, or Deceased.

Table B

Educational Classifications

9	Phd., Edd., LLd., M.D., DDS.
8	Master's Degree
7	B.A. or B. S. and more College
6	B.A. or B.S. only
5	Some College
4	High School Graduate
3	Some High School
2	Jr. High School Graduate
1	Elementary School Only

NEW MONTHLY DATA ON EMPLOYMENT AND UNEMPLOYMENT IN THE AMERICAN FAMILY

Howard Hayghe, Bureau of Labor Statistics

For almost 30 years, monthly statistics on employment and unemployment have been collected and published, providing such estimates as the number of employed adult men and women, the number of unemployed teenagers, or the number of wives entering or reentering the labor force. Over the long term, they have reflected the labor force effects of economic as well as social and demographic changes.

Today, these statistics have become an important part of the body of social indicators used in assessing the material well-being of American workers. The monthly data, however, have an analytical drawback in that they treat the work force as a group of discrete individuals, thereby touching only lightly and indirectly on the family dimension. The family, however, is still the primary economic unit in this country; most workers are family members--family heads, wives, children, or other relatives. Family members account for 88 percent of the 153 million men, women, and teenagers of working age (16 years and over) in the civilian noninstitutional population. 1/ What happens to these individuals in the labor market not only affects them, but their families as well.

In order to provide more relevant and timely information on families, the Bureau of Labor Statistics (BLS) is developing a new data series which will show the employment status of family heads, wives, or other relatives, crossed by the employment status of other members of their families. This new series, which is based on special tabulations from the Current Population Survey (CPS) and which begins with the first quarter of 1976, should provide particularly useful insights into the family situation of the unemployed. Specifically, it should tell us whether the unemployed were the only wage earners within their families, or whether other family members were employed thereby cushioning the effect of unemployment on the family unit.

During the first quarter of 1976, the number of unemployed persons, at 7.9 million (not seasonally adjusted) 2/, was only slightly below the recessionary peaks registered during 1975. Almost 6.9 million of the unemployed were members of families, including 2.5 million heads, over 1.5 million wives, and 2.8 million sons, daughters, or other relatives.

The new data give us a more complete picture. As shown in table 1, about 1.1 million, or two-fifths, of the families where the head was jobless there was at least one other family member employed. However, there were enormous differences between families headed by men and those headed by women. A far greater proportion of families headed by unemployed women had no other person employed than was the case for families headed by jobless men. This was due in large part to the fact that there was no one else of working age (16 years and over) in 59 percent of the families headed by unemployed women, compared with only some 2 percent of the male-headed families. Moreover, nearly half of the families headed by unemployed men had a wife or child who was employed.

This type of monthly data should be an important addition to the extensive information on families already available. For over a quarter of a century, data on families have been collected and published only on an annual basis (usually in March) through supplemental questions to the basic CPS. At BLS, the data are part of two series of Special Labor Force Reports: "Marital and Family Characteristics of Workers: beginning in 1959; joined by "Children of Working Mothers: in 1970. These publications include information on work experience of husbands and wives, family income, presence, age, and number of children, additional family workers, and much more. In addition to the labor force information published by BLS, household and family data from the March CPS are also available from the Bureau of the Census. Figures on the demographic characteristics of families, family income, as well as a host of other information, are published by Census in its series of Current Population Reports. The new series we are developing at BLS is designed to examine the effects of economic ups and downs on the employment situation in families on a quarterly basis, at first, but eventually on a monthly basis.

Besides giving a more complete view of unemployment, initial findings from the limited amount of preliminary new tabulations that are now available make the same outstanding point that has been evident in the once-a-year assessments of employment and unemployment in the family: Families where the head is employed are more likely to have other employed members than families where the head is either unemployed or not in the labor force. Thus, employment or lack of employment can be a "family affair."

There were a total of 56.3 million primary families 3/ in the first quarter of 1976. Of 41.8 million headed by an employed person, over half (53 percent) had at least one other member who was also employed. This proportion compares with 43 percent in families headed by an unemployed person, and only 29 percent in families headed by someone who was not in the labor force. These proportions and the differences between them vary a great deal by the characteristics of the families, such as whether they are male-headed or female-headed, black or white, and older or younger. However, the underlying relationship is remarkably consistent, as illustrated by the figures in table 2. To repeat, employment in the family is a "family affair," and conversely unemployment in the family is also a "family affair." Take wives for example.

Wives whose husbands were employed were more likely to be employed than were wives whose husbands were unemployed or not in the labor force. Of the women whose husbands had jobs in the first quarter of 1976, 17.3 million, or 47 percent also held jobs. However, as shown in table 3, when their husbands were unemployed, only 43 percent of the wives had jobs. Obviously wives were more likely to be unemployed if their husbands were also unemployed. In fact, historical trends from the annual supplemental family data of the March CPS indicate that the labor force participation rate for wives of unemployed husbands tends to be higher than that for wives of employed husbands. And, furthermore, these women have considerably more difficulty in finding jobs. For example, in the first quarter of 1976, the unemployment rate for wives of unemployed husbands--19.4 percent--was triple the rate for wives of employed husbands (6.5 percent). The magnitude of the differences between any of these rates for wives by the employment status of husbands has varied over time. Nevertheless, the pattern of differences has remained consistent over the long-term. Unemployment is unmistakably a "family affair."

Black wives were more likely to be working than white wives, (51 and 41 percent, respectively) whatever the employment status of their spouses. This is in line with the historically higher degree of labor force participation on the part of black than white wives. However, it should be noted that this gap has been narrowing. In March 1975, 54 percent of the black wives and 44 percent of the white were in the labor force, compared with 47 and 34 percent, respectively, 10 years earlier.

In addition to wives, there were about 23 million other relatives--primarily sons and daughters of whom 80 percent were under 25--in families headed by men. About 10.5 million of these other relatives were employed and two-fifths of these, including about half the employed 16 to 24-year-olds, worked part time. Like wives, these relatives were more likely to be working if the head were employed than if he were not. Where the head was employed, 47 percent of the other relatives were also employed, compared with about 43 percent where he was either unemployed or not in the labor force.

The unemployment rate of these other family members was quite high. As was the case with wives, the incidence of unemployment was higher for those in families where the head was jobless than where he was a jobholder. The unemployment rate of these other family members was 25 percent where the head was seeking work, compared with 16 percent where he was employed.

The only exception to the pattern for all families was among those headed by persons 65 years and over. The proportion of these older families with employed members was the same whether the head was employed or unemployed, probably because most members of these families are also older and hence likely to be out of the labor force.

An additional point worth noting is that families headed by someone 45 to 64 years old are far more likely than other families to have other employed members whatever the labor force status of the head. This is probably due, in part, to the fact that many of the children in these families are likely to have reached working age.

Families headed by women are a small but growing proportion of American families. In the first quarter of 1976, there were only about 7.4 million female-headed families compared to about 49 million headed by men. The labor force participation rate of female family heads, at 56 percent, was lower than for male heads (82 percent), and their unemployment rate was higher, 10 versus 5 percent. Also, about 1 out of 3 were living at or below the poverty level.

Because female family heads, by definition, have no spouse living with them, and because a high proportion (41 percent compared to less than 1 percent of the families headed by men) have no one else of working age, relatively fewer female-headed than male-headed families have other workers. In the first quarter of 1976, only 28 percent (2.1 million) of the female-headed families had someone employed excluding the head, compared with 50 percent of the male-headed families. Even so, a higher proportion of families where the head was employed than where she was not had additional workers. This was true for black families headed by women as well as white. However, blacks, whatever the employment status of the head, were less likely than white families to have additional workers. This was probably because relatively more black than white families headed by women had no other members of working age.

So far, this new family series is fairly limited in scope. Currently, we can examine monthly changes in the labor force status of family members, but as yet, we have no means of measuring the impact of economic events on family welfare. That is, data are not available on certain elements that would enable us to quantify this impact--elements such as duration of unemployment, size of family, and earnings. Much developmental work remains to be done to improve and broaden the usefulness of this data series. Also, until enough observations are available, it will be impossible to compute the seasonal component of any changes in employment and unemployment. 4/

Despite these drawbacks, the monthly data series on employment in the family can provide new and valuable insights into the family as an economic unit. With it we can observe some of the effects of economic trends on families and some of the reactions of family members to these events. These data will provide researchers with empirical tools to broaden their knowledge of family economics, as well as a timely factual base to evaluate policies and programs designed to soften the hardships engendered by economic fluctuations.

Footnotes

1/ The remaining 12 percent was composed of primary individuals, i.e., heads of households who have no other relatives living with them and other individuals who are not living with relatives, e.g., boarders.

2/ The seasonally adjusted number for the first quarter of 1976 was 7.2 million.

3/ A primary family includes the head of a household and all other persons in the household related to the head by blood, marriage, or adoption.

4/ The Bureau of Labor Statistics uses the Census X-11 method for seasonal adjustment of of data series. This is an adaptation of the standard ratio-to-moving average method with a provision for moving "adjustment factors" to take account of changing seasonal patterns. With this method, 36 observations (12 quarterly observations) are needed to seasonally adjust data.

Table 1. Unemployed family heads by employment status of other family members, first quarter, 1976

(Primary families 1/)

	Unemployed family heads		
	Total	Male	Female
Families with unemployed heads (number in thousands)	2,524	2,097	427
One or more other members employed 2/	1,092	1,021	71
Full time 3/	833	791	42
Part time only	259	230	29
One or more other members unemployed, none employed	282	238	44
No other members in labor force	1,150	838	312

1/ A primary family includes the head of a household and all other persons in the household related to the head by blood, marriage or adoption.

2/ In addition to employed persons, may also include some unemployed.

3/ May include some part time.

Table 2. Employment status of family members by employment status of family head, first quarter 1976

(Primary families 1/)

Employment status of family members	Employment status of family head		
	Employed, total	Unemployed, total	Not in labor force, total
ALL FAMILIES			
Total number (thousands)	41,774	2,524	12,002
Percent with....			
One or more other members employed 2/	53	43	29
Full time 3/	38	33	22
Part time only	15	10	7
White	53	44	29
Black	53	37	26
One or more other members unemployed, none employed	4	11	3
No other member in labor force	43	46	68
Head aged: 16 to 24 years	48	34	19
25 to 44 years	49	39	23
45 to 64 years	60	56	41
65 years and over	36	35	25
HUSBAND-WIFE FAMILIES			
Total number (thousands)	36,881	1,998	8,288
Percent with employed wives	47	43	21
White	46	43	20
Black	57	50	31
FEMALE-HEADED FAMILIES			
Total number (thousands)	3,688	427	3,280
Percent with....			
One or more other members employed 2/	30	17	27
White	32	21	32
Black	26	7	17
One or more other members unemployed, none employed	6	10	5
No other member in labor force	64	73	67

1/ See footnote 1, table 1.

2/ See footnote 2, table 1.

3/ See footnote 3, table 1.

Table 3. Employment status of husbands by employment status of wives, first quarter 1976

(Primary families 1/)

Employment status of wife	Employment status of husband		
	Employed	Unemployed	Not in labor force
Total: Numbers in thousands.......	36,881	1,998	8,288
Percent....................	100	100	100
Employed.............................	47	43	21
Unemployed...........................	3	10	2
Not in labor force...................	50	46	77

1/ See footnote 1, table 1.

DALLYING WITH SOME CPS DESIGN EFFECTS FOR PROPORTIONS

Thomas N. Herzog and Fritz Scheuren*
Social Security Administration

For large-scale complex surveys the cost of directly obtaining estimates of the variance of every statistic of interest may be prohibitive. It becomes necessary, therefore, to generalize from the variances one can afford to calculate to those one cannot. Practitioners have employed a number of ad hoc solutions to this extrapolation problem [e.g., 1-5]. The purpose of the present paper is to compare some of these proposed estimators. The scope of these comparisons, as the title of this note should suggest, is quite limited. In particular, the discussion will be confined to just a few of the issues concerning the estimation and averaging of design effects for proportions.

The paper is divided into two sections. In section 1 comparisons are made between the "usual" replicate estimator of a survey's design effect and an alternative suggested by Kullback's theory of minimum discrimination information [6]. The empirical work done indicates that these different estimators are virtually interchangeable. In section 2 we consider 14 distinct schemes for averaging design effects. Unlike our results in section 1 the type of average used seems to make a great deal of difference.

1. ESTIMATING DESIGN EFFECTS FOR PROPORTIONS

Often consistent estimators $\{\tilde{\sigma}_i^2\}$ of the simple random sampling variances $\{\sigma_i^2\}$ are readily available from a complex survey. This obviously applies to the case of proportions and may also apply to that of regression coefficients when these are calculated using standard computer packages [3]. If, in such situations, cost or space limitations make it impossible to provide direct estimates $\{\tilde{v}_i^2\}$ of the actual survey variances, a reasonable way to proceed is to examine what Kish [2] calls "design effects." These may be defined by the expression

$$(1) \qquad \delta = \frac{v^2}{\sigma^2}$$

where

v^2 is the actual (expected) survey variance.

σ^2 is the variance one would have obtained from a simple random sample (with replacement) of exactly the same size. In particular, for a proportion, p,

$$\sigma^2 = p(1-p)/n$$

where "n" is the total sample size.

δ is a measure of the variance impact of the complexities of the sample design relative to simple random sampling (that is, δ summarizes the combined effect of the number and nature of the selections at each stage of sampling, the extent of pre- and post-stratification, the ultimate cluster size, etc.).

The "Usual" Replicate Estimator of δ.--There are, of course, many ways to estimate design effects. Our purposes here would not be served by describing all of them. Instead, we will confine our attention to estimators of "δ" which pertain when the survey can be divided into "r" independent identically distributed subsamples or replicates. 1/

To fix ideas and to motivate the discussion which is to follow, consider the contingency table

TABLE 1.--Proportions of units with income under (over) \$10,000 by replicate

Dependency unit characteristic	Replicate 1	2	...	r
Income under \$10,000.....	$\tilde{p}_1$	$\tilde{p}_2$	...	$\tilde{p}_r$
Income of \$10,000 or more........	$(1-\tilde{p}_1)$	$(1-\tilde{p}_2)$	...	$(1-\tilde{p}_r)$

where for the j^{th} replicate, $j=1,\ldots,r$,

$\tilde{p}_j$ is an estimator of the proportion of units who reported income of less than \$10,000; and

$(1-\tilde{p}_j)$ is an estimator of the proportion who reported an income of \$10,000 or more.

The $\{\tilde{p}_j\}$ will be assumed to be independent, identically distributed, consistent estimators of the underlying (finite) population proportion p.

Now to estimate the design effect of the statistic

$$(2) \qquad \bar{p} = \frac{1}{r}\sum_{j=1}^{r} \tilde{p}_j$$

a common procedure is to calculate

$$(3) \qquad Y_1 = \frac{\sum_{j=1}^{r} (\tilde{p}_j-\bar{p})^2/r(r-1)}{\bar{p}(1-\bar{p})/n} .$$

This is because

$$(4) \qquad E\left\{\sum_{j=1}^{r} (\tilde{p}_j-\bar{p})^2/r(r-1)\right\} = v^2$$

and, for large n,

$$(5) \qquad E\left\{\bar{p}(1-\bar{p})/n\right\} \doteq \sigma^2.$$

Hence

$$(6) \qquad E\,Y_1 \doteq \delta.$$

Asymptotic Distribution of Y_1.--Under suitable regularity conditions, it can be established that Y_1 is distributed asymptotically as a constant times a chi-square random variable with (r-1) degrees of freedom. This is most easily seen by considering the Pearson X^2 statistic

$$(7)\quad X^2 = \sum_{j=1}^{r} \frac{n}{r}\left\{\frac{[\tilde{p}_j-\bar{p}]^2}{\bar{p}} + \frac{[(1-\tilde{p}_j)-(1-\bar{p})]^2}{(1-\bar{p})}\right\} .$$

$$= \sum_{j=1}^{r} \frac{(\tilde{p}_j-\bar{p})^2/r}{\bar{p}(1-\bar{p})/n} = (r-1)\ Y_1 .$$

Cramér [8] has shown that X^2 is asymptotically distributed as a chi-square with (r-1) degrees of freedom under fairly broad conditions provided one is engaged in simple random sampling and that the null hypothesis of homogeneity holds. Because the columns of table 1 are replicates the null (homogeneity) hypothesis is satisfied. Therefore, it only remains to examine the behavior of X^2 when the sampling design is other than simple random sampling. We will not do this in general; instead we will simply observe that if, as would be true for many complex designs,

(i) the $\tilde{p}_j$ are approximately normally distributed and
(ii) the effective sample size is sufficiently large so that $\bar{p}(1-\bar{p})$ can be treated as constant,

then

$$(8)\quad X^2 \sim \delta \left\{\chi^2(r-1)\right\} .$$

Hence

$$(9)\quad Y_1 \sim \frac{\delta}{(r-1)} \left\{\chi^2(r-1)\right\}$$

and the asymptotic variance of Y_1 is

$$(10)\quad V(Y_1) \doteq \left\{\frac{\delta}{(r-1)}\right\}^2 2(r-1) = \frac{2\delta^2}{(r-1)}$$

Other replicate estimators of δ.--There are a number of statistics which have the same asymptotic distribution as Y_1 but which might have better "small sample" properties. One such statistic is

$$(11)\quad Z_1 = \frac{1}{(r-1)}\ [2I(\tilde{p}_j:\bar{p})]$$

$$= \frac{2n}{r(r-1)} \sum_{j=1}^{r} \left\{\tilde{p}_j \ln [\tilde{p}_j/\bar{p}] + (1-\tilde{p}_j) \ln [(1-\tilde{p}_j)/(1-\bar{p})]\right\} .$$

Expression (11) is suggested by the relationship between the Pearson X^2 statistic and the Information-Theoretic approach to contingency tables [9].

Since neither Y_1 nor Z_1 is an unbiased estimator of δ (except asymptotically), it is reasonable to consider bias-reducing techniques such as jackknifing [10]; therefore, in the discussion of the empirical results which follows, comparisons have been made not only between the usual (Y_1) and information (Z_1) estimators but also between their corresponding jackknifed versions. These jackknifed statistics are

$$(12)\quad Y_2 = rY_1 - \left[\frac{r-1}{r}\right] \sum_{j=1}^{r} Y_1^{(h)}$$

and

$$(13)\quad Z_2 = rZ_1 - \left[\frac{r-1}{r}\right] \sum_{h=1}^{r} Z_1^{(h)}$$

where $Y_1^{(h)}$ (or $Z_1^{(h)}$) is computed in the same way as Y_1 (or Z_1), except that the h^{th} replicate is eliminated from the calculations and the number of replicates is taken as (r-1) not r.

Jackknifing also provides a (well-known) method for approximating the variance of all the estimators. For example, for Y_1 (or Y_2) the variance can be estimated by

$$(14)\quad \tilde{V}(Y_1) = \left\{\frac{r-1}{r}\right\}\left[\sum_{h=1}^{r}\left\{Y_1^{(h)}\right\}^2 - rY_3^2\right]$$

where

$$Y_3 = \frac{1}{r}\sum_{h=1}^{r} Y_1^{(h)} .$$

An equivalent expression for Z_1 (or Z_2) is

$$(15)\quad \tilde{V}(Z_1) = \left\{\frac{r-1}{r}\right\}\left[\sum_{h=1}^{r} Z_1^{(h)\,2} - rZ_3^2\right]$$

where

$$Z_3 = \frac{1}{r}\sum_{h=1}^{r} Z_1^{(h)} .$$

Background for Empirical Comparisons Made.--The standard procedure for looking at the small (finite) sample properties of estimators having identical asymptotic distributions is to conduct Monte Carlo experiments under varying, but carefully controlled, conditions. This approach has not been taken here. Instead, we will examine the estimators in the real-world context which originally motivated this paper.

Periodically, the Census Bureau, the Social Security Administration (SSA) and the Internal Revenue Service (IRS) have engaged in joint projects to study the reporting of income in the March Supplement to the Current Population Survey (CPS). One of these studies, which has been the subject of many papers at this and previous annual meetings [e.g., 11], was carried out using the March 1973 CPS and involved the linkage of survey schedules with administrative information from IRS and SSA records.

Coverage differences and matching difficulties must, of course, be addressed in such an exercise. Alternative adjustment strategies are often possible and issues concerning their impact on the sample variance arise [12]. When the assessment involves comparisons among a great many variances (as in this case [13]) summary descriptors, such as average design effects, appear attractive. The practical problems we faced, therefore, were how to estimate the design effects and how to combine or average them.

<u>Scope of empirical work and some limitations</u>.--We examined the design effects for CPS "Dependency Units" by race of the unit head. 2/ Within each racial group, percentage distributions of the units were produced separately for six different classifiers: type of unit, total size of unit, number of individuals 14 years of age or older in the unit, total earnings of unit members, total social security benefits received, and total unit income. Since the CPS is not composed of independent identically-designed subsamples [15], some practical compromises were necessary in our preliminary calculations. The "replicates" employed were the eight rotation panels in the March 1973 CPS.

Because the same sample of PSU's is common to all panels it is not possible to use the panels to estimate the between-PSU component of the CPS variance. Thus, an immediate consequence of this is that the "design effects" considered here relate only to the within-PSU variance component of the estimators. It might be mentioned parenthetically that for each statistic studied in this paper the within-PSU component probably accounts for 90 percent or more of the total variance.

Although all eight panels are initially selected by the Census Bureau in the same way, in any one survey month the individuals in each rotation group will have been interviewed a different number of times. Changes in the way interviews are conducted also occur during the life of a panel. Initially, the questions are asked in person; but, in the later panels, most of the surveying is done by telephone. The net effect of these and other factors [16] is to alter the response patterns from panel to panel so that the panels cannot be assumed to be identically distributed.

The influence of these panel differences on the statistics under consideration here is not known. 3/ When we began this work we implicitly assumed that such panel effects, if any, would be small enough to ignore. This was in part a reflection of our, perhaps misplaced, confidence in the nature of the raking ratio estimation procedures to be employed [17-19]. Project plans call for a repetition of the present calculations using a random group estimator (described in [20]) that would not be subject to "panel biases." These will be ready shortly and may be obtained on request. 4/

<u>Summary of Results</u>.--Space limitations do not permit us to show all the comparisons made among the four design effect estimators considered. Table 2 was created, therefore, to provide a summary description. A brief glance at it should convince the reader that, in general, the usual and information estimators, whether jackknifed or not, were virtually identical numerically for the statistics considered. Some other patterns we found of interest in the table are:

1. The jackknifing tends to slightly reduce the estimated design effect for both the usual and information statistics--less so in the latter case, however.

2. The information estimator tends to be slightly larger than the usual estimator.

Table 2.--Selected Comparisons among Alternative Within-PSU CPS Design Effect Estimators

(Mean values for each classifier of ratios shown. Underlying percentage distributions were obtained using Preferred Census Undercount "Corrected" Weight [19].)

Classifier 5/	Ratio Comparisons for White Units						Ratio Comparisons for Other Units					
	Estimators				Estimated Variances		Estimators				Estimated Variances	
	Original v. Jackknife		Usual v. Information		Relative Standard Errors	Relative Coefficients of Variation	Original v. Jackknife		Usual v. Information		Relative Standard Errors	Relative Coefficients of Variation
	Usual $Y_1 \div Y_2$	Information $Z_1 \div Z_2$	Original $Y_1 \div Z_1$	Jackknife $Y_2 \div Z_2$	$\left\{\tilde{V}(Y_1) \div \tilde{V}(Z_1)\right\}^{1/2}$	(5) ÷ (4)	Usual $Y_1 \div Y_2$	Information $Z_1 \div Z_2$	Original $Y_1 \div Z_1$	Jackknife $Y_2 : Z_2$	$\left\{\tilde{V}(Y_1) \div \tilde{V}(Z_1)\right\}^{1/2}$	(11) : (10)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Unit Type (6 classes)...........	0.996	0.999	1.005	1.008	1.011	1.003	0.999	1.000	0.997	0.999	1.014	1.015
Total Unit Size (10 classes)..........	1.001	1.001	0.996	0.995	1.006	1.011	1.005	1.002	0.987	0.985	0.997	1.012
Adults in Unit (7 classes)...........	1.000	1.000	0.997	0.997	1.014	1.017	1.016	1.006	0.965	0.956	0.975	1.021
Size of Total Earnings:												
Low (12 classes).......	1.001	1.001	0.997	0.996	0.998	1.002	1.000	1.000	0.997	0.998	0.997	1.000
High (10 classes)......	1.000	1.000	1.000	1.001	1.002	1.001	1.002	1.002	0.983	0.984	0.978	0.995
Size of Social Security Benefits (8 classes)...	1.001	1.001	0.996	0.996	0.996	1.000	1.000	1.000	0.999	0.999	0.974	0.975
Size of Total Income:												
Low (12 classes).......	1.000	1.000	1.000	1.000	1.001	1.001	1.003	1.001	0.993	0.991	1.006	1.016
High (11 classes)......	1.004	1.001	0.991	0.989	0.995	1.007	1.012	1.005	0.967	0.961	0.973	1.013

3. Both estimators have about the same estimated variances; the coefficient of variation of the information estimator is slightly smaller than that of the usual estimator.

4. Small differences exist in the above patterns by classifier. The differences by race are somewhat more important, however, and probably reflect the large disparity in the underlying sample sizes of the two groups.

All of these results, of course, are consistent with the asymptotic theory and suggest that in our application any of the estimators would be as suitable as any other.

Two further points need to be made lest misunderstandings arise. First, the results cast little or no light on the interpretative problems raised by the possibility of panel biases. Second, table 2 provides almost no information about the probability distribution of our estimators.

Table 3 below was constructed to look at this "distribution" question. The focus in this table is on the statistic

$$(16)\quad S = \left[\left\{\frac{2Y_2^2}{(r-1)}\right\} \div \left\{\tilde{V}(Y_1)\right\}\right]^{1/2}$$

which is a measure of the extent to which the variance estimator suggested by (10) agrees with the jackknife estimator (14).

As can be seen, much larger discrepancies exist in table 3 than in table 2. Some attempt to explain these seems necessary. One possible explanation is that the normality assumption required to obtain expression (10) does not hold exactly. To see that this idea could have merit, assume that all the required regularity conditions hold except normality; then, the standard error of Y_1 is approximately

Table 3.--Comparison by Race between Chi-Square and Jackknife Standard Error Estimates of Selected Within-PSU CPS Design Effects

Estimated Proportion Expressed as Min{p, (1-p)}	White Units		Other Units	
	Number of Comparisons	Mean of "S"	Number of Comparisons	Mean of "S"
Overall...	76	1.107	69	1.210
Less than 2%.	13	1.089	17	1.135
2 under 5%...	35	1.093	17	1.292
5 under 10%..	13	1.142	19	1.300
10 under 20%.	4	1.069	10	1.065
20 under 50%.	11	1.148	6	1.147

$$(17)\quad \text{S.E.}\ (Y_1) \doteq \delta\left\{\frac{2}{r-1} + \frac{\gamma_2}{r}\right\}^{1/2}, \text{ with } r=8,$$

where γ_2 is the customary measure [21] of the kurtosis of the $\{\tilde{p}_j\}$ and is zero under normality. It is fairly obvious that if γ_2 took on values which some might consider "close" to normality (roughly -0.2 for whites and -0.5 for other races), then at least the overall departures in table 3 could be reconciled. Other explanations, of course, can also be conjectured including possible differences in the distributions of the $\{\tilde{p}_j\}$ from panel to panel.

2. AVERAGING DESIGN EFFECTS FOR PROPORTIONS

In this section we will continue our analysis of the CPS data summarized in tables 2 and 3. Here our focus will shift from an examination of different estimators of the same design effect to a consideration of alternative schemes for averaging the same estimator (Y_1) over different design effects.

Averaging methods.--Four basic types of "averages" were explored: the median and three means (arithmetic, geometric, and harmonic). The averaging was done separately for whites (76 classes) and other races (69 classes). Three different weighting methods were studied: simple (unit) weighting of the estimated design effects, weighting by the inverse of the estimated simple random sampling (SRS) variances and weighting by the inverse of the estimated SRS relvariances. Results for these 4x3=12 schemes are shown in table 4. Also shown in that table are two other "averages":

1. Kish (square root) approach.--In [2] and [3], Kish recommends using the average of the square roots of the individual design effect estimates. Kish [2:p. 519] even prefers this to the unweighted arithmetic mean of the design effects.

2. Overall ratio average.--This is just the overall average of the replicate variances divided by the corresponding average of the simple random sampling variances.

Summary of results.--We did not expect that the 14 distinct averaging techniques considered here would produce the diverse numerical results displayed in table 4. Our first reaction, therefore, was that one or more programming errors had been made. So far, however, our checking indicates that programming errors are not the source of the differences.

The most plausible explanation for the large numerical discrepancies in table 4 is that the expected values of the estimated design effects being averaged are quite dissimilar. Since each of the schemes assigns somewhat different weights to the individual estimates, any lack of homogeneity might result in unequal expectations among the averages.

We have made some attempts at partitioning the design effects into subclasses which would be more homogeneous. In particular, the estimated

Table 4.--Summary Comparison among Alternative Within-PSU CPS Design Effect Averaging Proposals with Usual Estimator

Averaging Scheme	White Dependency Units				Other Dependency Units			
	Original	Jack-knifed	Standard Error	Coefficient of Variation (3) ÷ (2)	Original	Jack-knifed	Standard Error	Coefficient of Variation (7) ÷ (6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Uniform Unit Weighting:								
Arithmetic................	1.563	1.564	0.228	0.146	1.233	1.230	0.102	0.083
Geometric.................	1.372	1.371	0.203	0.148	0.987	0.983	0.087	0.089
Harmonic..................	1.192	1.189	0.205	0.172	0.733	0.728	0.152	0.209
Median....................	1.432	1.424	0.162	0.114	1.009	1.010	0.186	0.185
Weighting by Reciprocal of Estimated Variances under Simple Random Sample:								
Arithmetic................	1.325	1.341	0.258	0.193	1.203	1.201	0.140	0.117
Geometric.................	1.171	1.178	0.203	0.172	0.921	0.916	0.075	0.081
Harmonic..................	1.033	1.033	0.187	0.181	0.588	0.582	0.185	0.319
Median....................	1.069	1.093	0.176	0.161	0.979	0.972	0.255	0.263
Weighting by Reciprocal of Estimated Relvariance under Simple Random Sample:								
Arithmetic................	2.050	2.050	0.714	0.348	1.226	1.224	0.166	0.135
Geometric.................	1.804	1.803	0.599	0.332	1.020	1.018	0.191	0.188
Harmonic..................	1.550	1.549	0.505	0.326	0.861	0.859	0.208	0.243
Median....................	1.903	1.910	0.734	0.384	0.977	0.947	0.217	0.229
Kish Approach................	1.466	1.466	0.212	0.145	1.109	1.105	0.089	0.080
Overall Ratio Average........	1.917	1.917	0.590	0.307	1.252	1.250	0.143	0.114

design effects $\{Y_1\}$ were examined as a function of the estimated proportions $\{\bar{p}\}$. The range of the averages can be narrowed considerably by using three or four groupings which depend on $\bar{p}$. However, the numerical differences among the alternatives still remain "uncomfortably" large.

SOME CONCLUSIONS

There is no question that more than "dallying" may be required to "resolve" the issues raised here. We were encouraged by the equivalence of the Information-theoretic and "usual" estimators discussed in section 1. While there is a need for caution (as we indicated), the results were promising. The results of section 2, however, were not encouraging. The large numerical differences in the averages studied have given us pause and we expect to try other approaches in the near future. Moreover, we now have some doubts about the suitability of constructing summary design effects when comparing alternative adjustment techniques (our original purpose). It is possible, for example, that what are homogeneous groups for one adjustment procedure are not for another. Hence, comparisons between them could be quite sensitive to the way we happened to average the design effects.

FOOTNOTES

* The authors would like to thank Gary Shapiro of the Census Bureau for carefully examining an earlier draft of this paper, especially on issues surrounding the interpretation of the CPS calculations. Typing assistance was provided by Joan Reynolds.

1/ It should be noted that while pseudo-replicate variance estimation is not explicitly considered in this paper much that is said here can be readily generalized to deal satisfactorily with such procedures [e.g., 7].

2/ A dependency unit is a group of individuals in a CPS household who would generally be considered to be interdependent under social insurance programs [14].

3/ To the extent, however, that there are any panel differences these would lead to an increase in the expected value of the estimated design effects.

4/ The address is Social Security Administration, Office of Research and Statistics, 1875 Connecticut Ave., N.W., Washington, D.C. 20009.

5/ The number of classes shown for each classifier refers to the ratio comparisons made for white units. For other units, the number of classes used was slightly less. (Altogether, there were 76 classes for whites and 69 for other races.)

REFERENCES

[1] Hansen, M. H., Hurwitz, W. N. and Madow, W. G., Sample Survey Methods and Theory, Vol. I, New York, Wiley, 1953, pp. 571-577.

[2] Kish, L., Survey Sampling, New York, Wiley, 1965, pp. 574-582.

[3] Kish, L. and Frankel, M. R., "Balanced Repeated Replications for Standard Error," Journal of American Statistical Association, Vol. 65, 1970, pp. 1071-1094.

[4] Gonzalez, M., Ogus, J. L., Shapiro, G., and Tepping, B. J., "Standards for Discussion and Presentation of Errors in Survey and Census Data," Journal of American Statistical Association, Vol. 70, No. 351, part II, 1975, pp. 1-23.

[5] Finch, R. H., "Examples of Sampling Variability in the Continuous Work History Sample," CWHS Users' Conference, May 13, 1976.

[6] Kullback, S., Information Theory and Statistics, New York, Dover, 1959.

[7] McCarthy, P. J., Vital and Health Statistics, PHS Publication No. 1000, Series 2, Nos.14 and 31 (especially No. 31, pp. 12-15).

[8] Cramér, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946, pp. 446-450.

[9] Gokhale, D. V. and Kullback, S., The Information in Contingency Tables, to be published.

[10] Miller, R. G., "The Jackknife - a Review," Biometrika, Vol. 61, 1974, pp. 1-15.

[11] Kilss, B. and Alvey, W., "Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972," American Statistical Association Proceedings Social Statistics Section, 1977.

[12] Scheuren, F., and Oh, H. L., "Fiddling Around with Nonmatches and Mismatches," 1975 American Statistical Association Proceedings Social Statistics Section, 1976, pp. 627-633.

[13] Scheuren, F., "Adjusting for Undercoverage and Administrative Nonmatches - Preliminary Notes and Basic Table Specifications," SSA unpublished working paper, March 1976.

[14] Projector, D. S., et al, "Projection of March Current Population Survey: Population, Earnings, and Property Income, March 1972 to March 1976," Studies in Income Distribution, DHEW Publication (SSA) 75-11776, 1974.

[15] Thompson, M. H., and Shapiro, G., "The Current Population Survey: An Overview," Annals of Economic and Social Measurement, Vol. 2 1973, pp. 105-129.

[16] Bailar, B., "The Effects of Rotation Group Bias on Estimates from Panel Surveys," Journal of American Statistical Association, Vol. 70, 1975, pp. 23-30.

[17] Ireland, C. T. and Scheuren, F., "The Rake's Progress," Computer Programs for Contingency Table Analysis, Social Security Administration (Multilith), 1975, pp. 155-216.

[18] Vaughan, D. R. and Ireland, C. T., "Adjusting for Coverage Errors in the March 1973 Current Population Survey," 1975 American Statistical Association Proceedings Social Statistics Section, 1976, pp. 688-693.

[19] Scheuren, F. and Oh, H. L., "Initial Tables for Evaluating Adjustments for CPS Coverage Errors," SSA, unpublished memorandum, April, 1976.

[20] Bounpane, P., "General Outline of the Steps in the Keyfitz Estimate and Variance Programs," U.S. Census Bureau, unpublished memorandum, November, 1974.

[21] Scheffé, H., The Analysis of Variance, New York, Wiley, 1959, pp. 331-336.

AN ALGORITHM FOR ADJUSTING n-DIMENSIONAL TABULAR DATA TO CONFORM TO GENERAL LINEAR CONSTRAINTS

Kevin Hollenbeck, Mathematica Policy Research, Inc.*

1. Introduction

The purpose of this paper is to report progress on a computational algorithm being developed which adjusts tabular data to conform to linear constraints. A common application for which the algorithm is suited is the statistical adjustment of data to known marginal constraints. The particular problem which motivated this research, however, was the adjustment of data subject to the constraints that the ratios of some cell entries were to be kept constant. The paper provides, in turn, an example of a problem for which the algorithm was developed, the analytical formulation and solution of the general class of problems, and the computational experience gained to date in implementing the solution technique.

Briefly, the advantages of the proposed solution algorithm besides the fact that it allows general linear constraints on the cell entries and not just marginal constraints, are that it:

a) allows and adjusts data elements which are zero,

b) identifies redundant or inconsistent constraints, thus decreasing computer time and avoiding the possibility of nonconvergence,

c) weights the adjustment for each element,

d) gives an exact solution in a non-iterative process,

e) yields estimates which satisfy the constraints to the problem and may be useful as initial estimates in iterative procedures to derive non-least squares estimates.

2. An Example Problem

A cross-tabulation of data from the May, 1976 Current Population Survey (CPS) by the attributes of employment status and age, where employment status is classified into the mutually exclusive and exhaustive classes of employment, unemployment, not in the labor force, and noncivilian status is given in table 1.

It will be noted that the civilian labor force in May was comprised of 94.216 million persons and the (non-seasonally adjusted) unemployment rate was 6.69 percent.

Suppose that a projection of the same table to May, 1977 was desired. One of the major econometric models has forecast that in quarter 2 of 1977, the total noninstitutional population 16 and over will be 158.3 million, the civilian labor force will have 96.7 million members, the overall civilian unemployment rate will be 6.4 percent, the civilian unemployment rate of persons 16-19 will be 16.8 percent, and the civilian labor force participation rate of persons 16-19 will be 56.0 percent.[1] The problem is to adjust the table entries so that they conform to the five constraints imposed by the macroeconomic forecasts. If the entries in table 1 are labelled y_{ij}, $i=1,\ldots,4$; $j=1,\ldots,4$, then the five constraints can be written as

$$(1)\quad \sum_{i=1}^{4}\sum_{j=1}^{4} y_{ij} = 158{,}300$$

$$(2)\quad \sum_{i=1}^{4}\sum_{j=1}^{4} y_{ij} = 96{,}700$$

$$(3)\quad -.064\sum_{i=1}^{4} y_{i1} + .936\sum_{i=1}^{4} y_{i2} = 0$$

$$(4)\quad -.168\, y_{11} + .832\, y_{12} = 0$$

$$(5)\quad .44\, y_{11} + .44\, y_{12} - .56\, y_{13} = 0.$$

The solution then is to minimize a distance criterion function subject to (1) through (5).

3. General Problem Formulation and Solution

The general problem is set up and analytically solved in this section of the paper.

Notation -

D - an n-dimensional matrix of observed data in which each dimension represents an attribute of the data entities (such as age, race, sex, employment status, etc.) and in which each dimension is divided into mutually exclusive and exhaustive classes of the attribute values,

c_i; $i=1,\ldots,n$ - the number of classifications in dimension i; note that c_i need not be same for all i,

$q = \prod_{i=1}^{n} c_i$ - number of elements in D,

x - a (q x 1) vector of the elements of D reordered to form a vector,

y - a (q x 1) vector of the estimates which satisfy the constraints,

E - the n-dimensional matrix of estimates,

M,m - an (a x q) matrix and (a x 1) vector which express the a marginal constraints on the problem; note a may equal 0,

N,n - a (b x q) matrix and (b x 1) vector which express the b nonmarginal constraints; note b may equal 0,

TABLE 1

EMPLOYMENT STATUS IN MAY, 1976 OF THE NONINSTITUTIONAL POPULATION 16 YEARS OF AGE AND OVER, BY AGE
(numbers in thousands)

Age	Employment Status				
	Employed	Unemployed	Not in Labor Force	Noncivilian	Total
16 - 19	7,732	1,434	7,886	368	17,420
20 - 24	12,208	1,501	4,905	808	19,422
25 - 64	65,241	3,236	28,338	964	97,779
65 +	2,731	133	18,857	0	21,721
TOTAL	87,912	6,304	59,986	2,140	156,342

S - an n-dimensional matrix of weights for each element of D; these weights are "indicators" of the errors associated with the cell entries, and in most applications, are directly proportional to the cell frequency counts, and

$\hat{W}$ - a (q x q) diagonal matrix of weights associated with the elements of x (i.e., the diagonalization of the vector that results when the elements of S are reordered in the same fashion as the elements of D to form x).

There are several criteria of "closeness" which may be used to adjust the data. Ireland and Kullback [6] demonstrate that the method of iterative proportions suggested by Deming and Stephan [2] minimizes discrimination information. Stephan [8] developed an algorithm which provides a least squares solution. More recently, Feinberg and Holland [3] consider a Bayesian approach, Grizzle, Starmer, and Koch [4] assume a linear model on functions of the cell probabilities and employ least squares, and Brown and Muenz [1] propose a reduced mean square error estimation technique for two dimensional contingency table.

The present study has been limited to using a weighted least squares criterion to exploit the linearity of the gradient of the objective function. (Exploration of the question of whether the analytical and computational solution techniques are applicable to other criteria is intended). The weighted least squares problem is:

$$(6) \quad \min_{e} \sum_{i=1}^{c_1} \sum_{j=1}^{c_2} \cdots \sum_{m=1}^{c_n} s_{ij\ldots m}(e_{ij\ldots m} - d_{ij\ldots m})^2 ,$$

where $s_{ij\ldots m}$, $e_{ij\ldots m}$, $d_{ij\ldots m}$ are typical elements of S, E, and D, and subject to a marginal constraints and b nonmarginal constraints.

Reordering the elements of D and E into vectors x and y and reordering the elements of S into the diagonal matrix $\hat{W}$, the problem can be stated as:

$$(7) \quad \min_{y} (y - x)^T \hat{W} (y - x)$$

$$\text{s. t. } M y = m$$
$$N y = n.$$

Analytically the problem is solved using straigntforward constrained maximization techniques. Form the Lagrangean function L:

$$(8) \quad L = (y - x)^T \hat{W} (y - x) - \Lambda_1 (m - My) - \Lambda_2 (n - Ny).$$

The first order conditions for a solution lead to the following set of simultaneous equations:

$$\begin{matrix}(9)\\(10)\\(11)\end{matrix} \begin{pmatrix} \hat{W} & M^T & N^T \\ M & 0 & 0 \\ N & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ \Lambda_1 \\ \Lambda_2 \end{pmatrix} = \begin{pmatrix} \hat{W}x \\ m \\ n \end{pmatrix} .$$

The second order conditions for a minimum are satisfied as long as the elements of $\hat{W}$ are positive. Premultiplying (9) by $M\hat{W}^{-1}$ and subtracting the result from (10) and premultiplying (9) by $N\hat{W}^{-1}$ and subtracting the result from (11) and then multiplying the Λ_1 and Λ_2 vectors by -1 yields the system:

$$
\begin{matrix} (9') \\ (10') \\ (11') \end{matrix}
\begin{pmatrix} I & -\hat{W}^{-1}M^T & -\hat{W}^{-1}N^T \\ 0 & M\hat{W}^{-1}M^T & M\hat{W}^{-1}N^T \\ 0 & N\hat{W}^{-1}M^T & N\hat{W}^{-1}N^T \end{pmatrix}
\begin{pmatrix} y \\ \Lambda_1 \\ \Lambda_2 \end{pmatrix}
=
\begin{pmatrix} x \\ m-mX \\ n-nX \end{pmatrix}.
$$

To solve the problem, it suffices to solve the subsystem of equations (10') and (11'). This is done by Gaussian elimination on the (a+b) x (a+b) symmetric matrix

$$
\begin{pmatrix} M\hat{W}^{-1}M^T & M\hat{W}^{-1}N^T \\ N\hat{W}^{-1}M^T & N\hat{W}^{-1}N^T \end{pmatrix}.
$$

By back substitution, we derive estimates of y:

$$(12)\quad y = \hat{W}^{-1}M^T\Lambda_1 + \hat{W}^{-1}N^T\Lambda_2 + x.$$

A very nice feature of this algorithm is that if the Gaussian elimination procedure to solve the subsystem of equations (10') and (11') cannot eliminate a row while it is pivotting because all of the elements are zero, then this implies that the associated constraint is redundant or inconsistent and that constraint is eliminated from consideration. This decreases computer time and may point out inconsistencies to the analyst.

4. Computational Experience

A program to solve the data adjustment problem using the above technique has been written for an IBM 370 system.[2] The program, LFNJUST,[3] was used to solve the above example problem. The data for the problem are given in table 2. The solution is given in table 3. It can be seen in that table that, except for rounding error, the projected population is 158.3 million; the civilian labor force is 96.7 million; the civilian unemployment rate is 6,187/96,699.8 = 6.40 percent; the civilian unemployment rate of persons 16-19 is 1,610.7/9,597.4 - 16.80 percent; and the civilian labor force participation rate of persons 16-19 is 9,587.4/17,120.4 = 56.0 percent. Thus all the constraints are satisfied.

5. Conclusions

The LFNJUST algorithm and program have demonstrated the utility of directly using constrained optimization techniques rather than iterative algorithms for the adjustment of tabular data to conform to linear constraints. Because it adjusts cell entries rather than frequencies, it is particularly useful when reweighting survey data. Additional study is warranted to consider whether the technique is efficient when minimum discrimination information estimates are desired and to consider multivariate table adjustment.

TABLE 2

DATA FOR EXAMPLE PROBLEM

$$
X^T = (7732 \;\; 12208 \;\; 65241 \;\; 2731 \;\; 1434 \;\; 1501 \;\; 3236 \;\; 133 \;\; 7886 \;\; 4905 \;\; 28338 \;\; 18857 \;\; 368 \;\; 808 \;\; 964 \;\; 0)
$$

$$
\hat{W}^{-1} = X = \begin{pmatrix} 7732 & & & & \bigcirc \\ & 12208 & & & \\ & & \ddots & & \\ & \bigcirc & & 964 & \\ & & & & 0 \end{pmatrix}
$$

$$
M = \begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0 \end{pmatrix}
$$

$$
N = \begin{pmatrix} -.064 & -.064 & -.064 & -.064 & -.936 & -.936 & -.936 & -.936 & 0&0&0&0&0&0&0&0 \\ -.168 & 0 & 0 & 0 & .832 & 0 & 0 & 0 & 0&0&0&0&0&0&0&0 \\ .440 & 0 & 0 & 0 & .440 & 0 & 0 & 0 & -.560&0&0&0&0&0&0&0 \end{pmatrix}
$$

$$
m = \begin{pmatrix} 158{,}300 \\ 96{,}700 \end{pmatrix}
$$

$$
n = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
$$

TABLE 3

ESTIMATED EMPLOYMENT STATUS IN MAY, 1977 OF THE NONINSTITUTIONAL POPULATION OVER 15 YEARS OF AGE, BY AGE (Numbers in thousands)

Age	Employment Status				
	Employed	Unemployed	Not in Labor Force	Noncivilian	Total
16 - 19	7,976.7	1,610.7	7,533.0	366.8	17,487.2
20 - 24	12,566.5	1,411.0	4,889.4	805.4	19,672.3
25 - 64	67,156.7	3,042.0	28,247.6	960.9	99,407.2
65 +	2,811.2	125.0	18,796.9	0.0	21,733.1
TOTAL	90,511.1	6,188.7	59,466.9	2,133.1	158,299.8

FOOTNOTES

*The research on which this paper is based was supported by the Socioeconomic Impact Division, Office of Economic Impact, Federal Energy Administration. Computational assistance was provided by Marjorie Odle.

1. The estimates were computed from the Data Resources, Inc. (DRI) CONTROL0524 forecast published in The Data Resources Review, June 1976 (Lexington, Mass.: Data Resources, Inc.).

2. The program was written by Marjorie Odle of The Hendrickson Croporation and is in the testing stage. It will be fully documented and available from the author on request after December 1, 1976. The program utilizes GELS, an IBM-developed procedure to solve a system of linear equations, in which the coefficient matrix is symmetric. See IBM [5].

3. The name is derived from its predecessor, NJUST, a least squares iterative adjustment algorithm developed by the Office of Research and Statistics, Social Security Administration, (see Pugh, Tyler, and George [7]), and because it will be applied to labor force adjustments.

REFERENCES

Brown, C. C. and Muenz, L. R., "Reduced Mean Square Error Estimation in Contingency Tables," Journal of the American Statistical Association, 71 (March 1976), 176-182.

Deming, W. E. and Stephan, F. F., "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known," Annals of Mathematical Statistics, 11 (1940), 427-44.

Feinberg, S. E. and Holland, P. W., "Simultaneous Estimation of Multinomial Cell Probabilities," Journal of the American Statistical Association, 68 (September 1973), 683-91.

Grizzle, J. E., Starmer, C.F., and Koch, G. G., "Analysis of Categorical Data by Linear Models," Biometrics, 25 (September 1969), 489-504.

International Business Machines, Inc., System/360 Scientific Subroutine Package Version III (White Plains, New York: IBM, Inc., 1968).

Ireland, C. T. and Kullback, S., "Contingency Tables with Given Marginals," Biometrics, 55 (March 1968), 179-88.

Pugh, R. E., Tyler, B. S., George, S., "Computer-Based Procedure for the n-Dimensional Adjustment of Data--NJUST," Unpublished Paper, Office of Research and Statistics, Social Security Administration (May 1974).

Stephan, F. F., "Iterative Method of Adjusting Sample Frequency Tables When Expected Margins are Known," Annals of Mathematical Statistics, 13 (1942), 166-78.

METHODS OF VALIDATING SELF-REPORTS OF DRUG USE: A CRITICAL REVIEW

Robert L. Hubbard, William C. Eckerman, and J. Valley Rachal
Research Triangle Institute

One of the most difficult tasks facing social researchers is the accurate assessment of an individual's nonmedical drug use. The usual problems of recall and recognition are compounded by the possible consequences of revealing socially undesirable and, often, criminal behavior. Some researchers estimate that the actual rates of nonmedical experience with certain drugs or recent use of illicit substances by respondents may be double the figures reported [5]. The evidence available to assess this claim is neither complete nor conclusive [5,12,25]. The few methodological studies that have been conducted have usually focused on a particular segment of the population such as addicts in treatment, drug users with arrest records or in jail, students, or servicemen. Many researchers check self-reports against a single criterion, which may also have questionable validity. Despite such limitations reviewers of these studies have concluded that the overall validity self-reports are acceptable for many research purposes. However, the problem is not only to establish an adequate level of overall validity in the general population but to determine the differential validity of reports in special segments of the population, among different patterns and types of drug usage, and under certain interviewing or questioning situations.

The purpose of this paper is to classify possible methods of validation to facilitate more systematic examination of the validity of self-reports of nonmedical drug use. These methods can generally be grouped into three categories: the examination of internal validity, an assessment of construct validity, and a determination of empirical validity.

INTERNAL VALIDITY

Internal validity can be indicated by rates of response to individual drug questions, the consistency of responses within an instrument, consistency of reports over time, and evidence of the candidness of the respondent.

Response Rates

Most studies report generally low refusal rates on individual items. Only 2.5 percent of the items that asked directly about use in a national sample of male high school students were not answered [16], and 3 percent in a study of enlisted men in Vietnam [23]. On the other hand, one paper reported that one of every eight college student respondents failed to complete a direct question on drug use, and one in 20 did not answer a more indirect question [13]. In another high school study, nonresponse rates to items on specific drugs varied from 1 to 6 percent [25].

Concurrent Consistency

The available evidence indicates that the responses to similar drug items within the same questionnaire are consistent. The greatest inconsistency between responses for a variety of products was 4 percent for usage of an "ups" drug classification [15]. Another study reported that only 15 of over 8,000 college student questionnaires contained "flagrant inconsistencies" [14]. Less than 1 percent of respondents answered questions on marihuana use differently compared to 2 percent on questions of alcohol usage. In a final study, few inconsistent responses were given to questions on ever trying, most recent use of, and age at first use of marihuana [25].

Another check on response consistency is the comparison of individual and aggregate rates of reported use under different techniques of eliciting drug use information. A number of studies have been unable to demonstrate that settings or instructions emphasizing anonymity or confidentiality have an impact on rates of reported use [15,17,18]. However, there is some evidence that greater frequency of use and more recent use is reported with anonymous procedures [18]. The design of the drug questions themselves also appears to affect rates of reported use. A randomized response technique guaranteeing anonymity produced fewer response refusals and higher estimates of drug usage among college students compared to a traditional questionnaire [13]. Among Army personnel differential rates of drug usage were reported in an anonymous questionnaire and in a randomized response technique as a function of the rank of the respondent [6]. These results suggest that different techniques may produce differing rates of response for respondents with varying levels of sophistication, education, and experience.

Consistency Over Time

Varying stabilities in responses over time have been found. Identical rates of reported use were found in questionnaires administered to high school students 2 weeks apart [15]. In a study of college students responding to the identical questionnaire 1 month later, 5 percent changed their answers [14]. The direction of change was equally distributed among those who changed their answers. In another study test-retest coefficients for reports of drug use ranged from 65 percent to 95 percent for patients transferring from one treatment facility to another [28]. Inconsistency was found to vary greatly depending on the type, pattern, and recency of use in a sample of high school students. Among illicit drug users, 9 percent of the marihuana users gave inconsistent responses over a 6-month period compared to 56 percent of illicit users of tranquilizers [25].

Candidness

A third method of establishing the internal validity of response patterns is by attempting to determine the honesty of the respondent. One direct method is to ask the respondents how honest they were. In a national survey, 78 percent of the adults interviewed reported that they were completely honest in their answers to questions on marihuana usage and did not hold back any information. Two percent later said they had had more experience, and 2 percent admitted less experience with marihuana than they had indicated on the questionnaire [2]. Another researcher found that 80 percent of the high school students in one study and 70 percent in a second "felt free to answer all the questions on drugs honestly" [15]. However, only 40 percent of the students felt that other students answered the drug questions honestly. In a study of enlisted men in Vietnam, 3 percent of the respondents reported that their answers were not completely honest [23]. In a longitudinal study 10 percent of the college students disclosed that they did not candidly answer questions on drug usage in the initial interview [14]. The substantial increase in the admission of "ever using" drugs in a followup interview of these students 1 year later could indicate an increased trust of the researchers, more social acceptance of use, or more peer pressure to identify oneself as a drug user rather than actual increases in use.

Another check on honesty in responding is the interviewer's perception of the respondent. Interviewers in a national study of marihuana use felt that 85 percent of the respondents were cooperative, and they were completely confident in 75 percent of the reports of use by the respondents [2]. These figures indicate that up to one in every five respondents may distort their drug usage. The impact of such distortions on the results of studies has rarely been examined.

Fictitious products have often been used to identify respondents who exaggerate use [15,20, 25,28]. However, the low proportion of reports for these products and their relationship to other use patterns seems to show that instead of an adequate check on accuracy of reporting, these techniques are more likely to identify respondents who are not familiar with particular products.

The results of an early validation study indicate less candor. A comparison of questionnaire admissions of using or selling narcotic drugs with subsequent responses with a polygraph showed only 10 percent of the respondents truthfully admitted either using or selling narcotic drugs on the questionnaire [8]. An additional 5 percent were detected as reporting falsely on the questionnaire. Two and a half percent of the respondents refused to answer this question. Based on these results, claims that actual use rates may be double those reported in studies in the late 1960's or early 1970's may not be exaggerated.

CONSTRUCT VALIDITY

A number of other procedures have been suggested that emphasize the correlation of use reports with other variables known or believed to be associated with drug use. The attribution of drug usage through the use of highly correlated variables avoids problems of probing sensitive areas, invading privacy, or identifying dishonest respondents.

One indirect method is the comparison of proportions of respondents admitting use with estimates of use by respondents or the proportions of use reported in other studies [12,16]. A second method involves the correspondence between self-reported use and use by friends [15, 16,25,28]. In one study 63 percent of the best friends of adolescent drug users also reported drug use compared to 22 percent of those of nonusers [25].

Users have been found to differ fron nonusers on a number of variables not directly connected to drug use. In one sample of high school students, absenteeism for users was triple that for nonusers [25]. In another, admitted drug use was related to attitudes toward use and behaviors thought to be associated with use including opposition to the Vietnam war, poor school performance, delinquency, and countercultural lifestyle [16]. Although a variety of demographic characteristics, personality traits, attitudinal variables, and sociocultural lifestyles have been found to be associated with drug use, correlation is not sufficient evidence of validity. Construct validation procedures may provide a good indication of the accuracy of use reports among stereotypic drug users. However, occasional users or users who do not have the characteristics of "typical" drug users may deny use of drugs as well as other forms of deviant or delinquent behavior.

EMPIRICAL VALIDITY

A direct check on the accuracy of self-reports is information on use from other sources. Official records and urinalyses are the principal criteria that have been employed.

Records

A landmark methodological study in the drug field used prescription records as the criterion. Validity coefficients for self-reports of medical drug use in the past year ranged from 83 percent for tranquilizers to as low as 64 percent for antibiotics [19]. These results do not compare unfavorably to the accuracy of reports of other information, such as voter registration (82 percent), library card ownership (87 percent) and age according to driver's license records (92 percent) [7].

It might be expected that the validity of reports of illicit drug use would be somewhat lower than reports of medical use of prescription drugs. Furthermore, it is difficult to find accurate records of illicit drug use. A recent attempt to validate reports of illicit drug use against treatment records suggested that treatment records contain many inaccuracies [1]. Drug registers or arrest records for narcotics offenses identify few users who are not identified through self-reports or urinalysis [11,24]. In a comprehensive study of the utility of using records for the identification of drug users, 44 percent of 190 respondents reported taking drugs and 3 percent reported addiction, although they had no police, medical, or armed service record of use [22].

Double-blind validation studies using records are a valuable and necessary part of any research effort to accurately assess the validity of self-reports. However, the substantial number of self-admitted users not listed in treatment records, drug registers, or arrest records indicate that such criteria have limitations.

Urinalysis

Current technological advances have made it possible to check the use of drugs by chemical methods. Urine samples taken at the time of a followup interview of returning Vietnam servicemen corresponded closely to the self-reports [21]. High correlations between self-reports and urinalysis results have been found for groups of known heroin or narcotic addicts [4,9]. However, two studies of arrestee populations demonstrate that self-reports and urinalysis corresponded for only 35 percent and 25 percent of respondents identified, respectively, as current narcotic or heroin users [11,24]. Almost half of the current narcotic or heroin users in both studies were identified by self-reports alone. In one of the studies 27 percent of the respondents with evidence of heroin in the urinalysis denied current usage. When the validity of reports of nonopiate use was also examined, the correspondence between interview data and urinalysis was much lower for amphetamine (30 percent) and barbiturate (33 percent) use than for heroin use (83 percent) [11]. These results suggest that the validity coefficients may be drug specific.

Chemical means of detection have a number of shortcomings. Only a limited number of drug substances are amenable to detection. Usual methods can only detect heroin (metabolized as morphine) in the system up to 24 hours after injection [10]. Use of a number of drugs cannot now be readily and economically detected, nor can medical use of legally prescribed drugs be distinguished from nonmedical use. Chemical analyses employed in urinalyses are subject to frequent technical breakdowns. Careful monitoring through the introduction of standard samples is necessary, particularly to maintain precision in the detection of amphetamines [11]. With such procedures and multiple tests, error rates of 1.5 percent can be obtained [3], which indicate much greater accuracy than self-reports.

Observation

The use of informants or participant-observers may have some utility in identifying drug users and, perhaps, the type, quantity and purity of drug use among various populations of users. One study found a high agreement of self-reports with information obtained from after-care counselors and relatives [27]. In such studies there may be legal and ethical questions at stake and the potential problem of jeopardizing respondent-researcher or client-patient relationships.

CONCLUSIONS

Accurate assessment of drug use by self-reports is fundamental to a variety of research efforts in the drug field. Estimates of the extent of untreated drug abuse, evaluations of treatment programs, and analysis of the relationship between drug use and its assumed consequences such as criminal behavior require valid information on drug use. Acceptable levels of validity need to be established for studies focusing on the prevalence of use at a particular time, trends in usage patterns over time, and relationships between drug usage and other behaviors.

Different levels of validity may be required for each, and certain aspects of validity should be emphasized, depending on the information needed. Some level of inaccuracy may be tolerable in estimates of prevalence because systematic reporting error can be incorporated into the calculation of confidence intervals. An acceptable level of such error should be specified, and appropriate guidelines for correction procedures should be established. Differential reporting accuracy by drug and respondent characteristics also needs to be determined.

Changes in rates of drug usage over time need to be separated from systematic measurement error. Trend estimates may be distorted by changes in validity of reporting over time due to increased social acceptance of some drugs as well as instability in response patterns. Such estimates require more precision than simple point prevalence estimates.

Maximum validity is required for the determination of the relationship between drug use and other behaviors, especially those which may also be underreported. For example, in one study, arrestees with drug-positive urines who denied use were more likely to have been arrested for crimes against persons than respondents who admitted use. By combining self-reports with urinalysis results, a more complete and accurate analysis of the relationship between drug use and criminal behavior was possible [11].

Careful validation and specified levels of validity minimize the misinterpretation of results due to inaccurate or dishonest reporting. As has been pointed out no single method or criterion is adequate to test the validity of self-reports. Combinations of internal, construct, and empirical validation procedures should be employed. Where feasible a variety of techniques should be systematically incorporated into the design of all studies where self-reports are the principal measure of drug use to determine if the measures have sufficient validity to accomplish the stated purpose of the study.

REFERENCES

[1] Abelson, H.J. and Atkinson, R.B. *Public Experience with Psycho-Active Substances.* Princeton, N.J.: Response Analysis Corporation, 1975.

[2] ______, Cohen, R., and Schrayer, D. "Public Attitudes Toward Marihuana, Part I: Main Report." In National Commission on Marihuana and Drug Abuse, *Marihuana: A Signal of Misunderstanding,* Appendix, Vol. II. Washington, D.C.: U.S. Government Printing Office, 1972.

[3] Baker, S.L., Jr. "U.S. Army Heroin Abuse Identification Program in Vietnam: Implications for a Methadone Program," *American Journal of Public Health,* 62 (1972) 857-60.

[4] Ball, J.C. "The Reliability and Validity of Interview Data Obtained from 59 Narcotic

Drug Addicts," *American Journal of Sociology*, 72 (1967), 650-4.

[5] Berg, D.F. "The Non-Medical Use of Dangerous Drugs in the United States: A Comprehensive View," *The International Journal of the Addictions*, 5 (1970), 777-834.

[6] Brown, G.H. and Harding, J.D. *A Comparison of Methods of Studying Illicit Drug Usage.* Alexandria, Va.: Human Resources Research Organization, 1973.

[7] Cahalan, D. "Correlates of Respondent Accuracy in the Denver Validity Study," *Public Opinion Quarterly*, 32 (1968-9), 588-633.

[8] Clark, J.P. and Tifft, L.L. "Polograph and Interview Validation of Self-Reported Deviant Behavior," *American Sociological Review*, 31 (1966), 516-23.

[9] Cox, T.J. and Longwell, B. "Reliability of Interview Data Concerning Current Heroin Use from Heroin Addicts on Methadone," *The International Journal of the Addictions*, 9 (1974), 161-5.

[10] Dole, N.P., Kim, W.K. and Eglitis, I. "Detection of Narcotic Drugs, Tranquilizers, Amphetamines and Barbiturates in Urine," *Journal of the American Medical Association*, 198 (1966), 349-52.

[11] Eckerman, W.C., Rachal, J.V., Hubbard, R.L., and Poole, K.W. "Methodological Issues in Identifying Drug Users." *Drug Use and Crime*, Rockville, Md., National Institute on Drug Abuse, (in press).

[12] Glenn, W.A. and Richards, L.G. *Recent Surveys of Nonmedical Drug Use: A Compendium of Abstracts.* Rockville, Md.: National Institute on Drug Abuse, 1974.

[13] Goodstadt, M.S. and Gruson, V. "The Randomized Response Technique; A Test on Drug Use," *Journal of the American Statistical Association*, 70 (1975), 814-8.

[14] Groves, W.E. "Patterns of College Student Drug Use and Lifestyles. In E. Josephson and E. E. Carroll (Eds.) *Drug Use: Epidemiological and Sociological Approaches.* Washington, D.C.: Hemisphere Publishing Corp. 1974, 241-75.

[15] Haberman, P.W., Josephson, E., Zanes, A., and Elinson, J. "High School Drug Behavior: A Methodological Report on Pilot Studies." In S. Einstein and S. Allen (Eds.) *Proceedings of the First International Conference on Student Drug Surveys.* Farmingdale, N.Y.: Baywood Publishing Co., 1972, 103-21.

[16] Johnston, L.D. "Drug Use During and After High School: Results of a National Longitudinal Study," *The American Journal of Public Health; Supplement*, 64, Part Two, (1974), 29-37.

[17] King, F.W. "Anonymous Versus Identifiable Questionnaires in Drug Usage Surveys," *American Psychologist*, 25 (1970), 982-5.

[18] Leutgert, M.J. and Armstrong, A.H. "Methodological Issues in Drug Usage Surveys: Anonymity, Recency and Frequency," *The International Journal of the Addictions*, 8 (1973), 683-9.

[19] Parry, H.J., Balter, M.B. and Cisin, I.H. "Primary Levels of Underreporting Psychotrophic Drug Use," *Public Opinion Quarterly*, 34 (1971), 582-92.

[20] Petzel, J.P., Johnson, J.E. and McKillip, J. "Response Bias in Drug Surveys," *Journal of Consulting and Clinical Psychology*, 40 (1973), 437-9.

[21] Robins, L.N. *A Follow-up of Vietnam Drug Users*, Washington, D.C.: Special Action Office for Drug Abuse Prevention, 1973.

[22] ______, and Murphy, G.E. "Drug Use in a Normal Population of Young Negro Men," *American Journal of Public Health*, 57 (1967), 1580-96.

[23] Roffman, R.A., and Sapol, E. "Marihuana in Vietnam: A Survey of Use Among Army Enlisted Men in the Two Southern Corps," *The International Journal of the Addictions*, 5 (1970), 1-42.

[24] Savitz, L. "Drug Use and Drug Users in An Arrestee Population." Technical Report, Philadelphia TASC Mass Urine Screening Program, 1974.

[25] Single, E., Kandel, D., and Johnson, B.D. *The Reliability and Validity of Drug Use Responses in a Large Scale Longitudinal Survey*, 5 (1975), 426-43.

[26] Smart, R.G.W. "Recent Studies of the Validity and Reliability of Self-Reported Drug Use, 1970-4," *Canadian Journal of Criminology and Corrections*, 17 (1975), 326-33.

[27] Stephens, R. "The Truthfulness of Addict Respondents in Research Projects." *The International Journal of the Addictions*, 7 (1972), 549-58.

[28] Whitehead, P.C. and Smart, R.G. "Validity and Reliability of Self-Reported Drug Use," *Canadian Journal of Criminology and Corrections*, 14 (1972), 1-7.

MULTIPLE COMPARISON TESTS FOR CONTRASTS AMONG CORRELATED CORRELATION COEFFICENTS: THE MULTIVARIATE PREDICTOR CASE

Bradley E. Huitema, Western Michigan University

Multiple comparison procedures for testing differences among correlation coefficients have recently been proposed for the independent sample case (Marascuilo, 1966; Huitema, 1974) as well as the correlated sample case (Huitema, 1975). A disadvantage of the procedures proposed for the latter case is that they are all conditional tests. The results of these tests generalize to a population in which the values of the predictor variables are fixed at the values included in the sample. In most correlation research interest lies in the generalization of results to future observed values of the predictor variables. That is, we wish to generalize the results to the whole range of predictor values included in the population from which the sample values were randomly selected rather than to a population including only the predictor values actually included in the sample. Methods of dealing with this problem in the case of two predictor variables have been proposed. The purpose of this paper is to suggest two methods of dealing with the multivariate normal case in which three or more predictors are involved and multiple comparisons among the correlated correlation coefficients are required.

Method I

Method I is proposed for tests on all pairwise comparisons. This method involves two stages:

Stage A: The overall hypothesis

$$H_0: \rho_{yx_1} = \rho_{yx_2} = \cdots = \rho_{yx_m}$$

(where ρ is the population correlation coefficient, y is the dependent variable and x_1 through x_m are the predictor variables) is tested using a generalization of Hotelling's well known test of the equality of two correlated correlation coefficients (Hotelling, 1940). If this test is significant at the specified level of alpha, proceed to stage B.

Stage B: Compute Williams' modification of Hotelling's two predictor test for each pairwise comparison using the same alpha employed during Stage A.

Computation Procedure

The computation procedure for Stages A and B are described in this section.

Stage A

Step 1

Compute $\underline{R}_x$, the interrcorrelation matrix of all predictor variables.

Step 2

Compute $\underline{R}^{-1}_x$, the inverse of $\underline{R}_x$.

Step 3

Compute $\underline{r}_{yx}$, the column vector of correlations between the dependent variable and each predictor variable, i.e.,

$$\underline{r}_{yx} = \begin{bmatrix} r_{yx_1} \\ r_{yx_2} \\ \cdot \\ \cdot \\ \cdot \\ r_{yx_m} \end{bmatrix}$$

Step 4

Compute the coefficient of multiple determination $R^2_{yx_1x_2 \cdots x_m}$.

Step 5

Obtain $\sum\sum c_{ij}$, the sum of the elements of $\underline{R}^{-1}_x$.

Step 6

Obtain $c_1, c_2, \ldots c_m$, the sums of rows 1 through m of $\underline{R}^{-1}_x$.

Step 7

Compute the weights $w_1, w_2, \ldots w_m$ where the ith weight is

$$w_i = \frac{c_i}{\sum\sum c_{ij}}.$$

The sum of the $w_i = 1$. The column vector of weights is denoted $\underline{w}$.

Step 8

Compute $\underline{w}' \, \underline{r}_{yx} = h$.

Step 9

Compute the product $h^2(\sum\sum c_{ij})$.

Step 10

Compute the F statistic using

$$F = \frac{\left[R^2_{yx_1x_2 \ldots x_m} - h^2(\sum\sum c_{ij})\right] / m-1}{(1 - R^2_{yx_1x_2 \ldots x_m})/N-m-1}.$$

The obtained F is evaluated with $F_{\alpha, m-1, N-m-1}$ where N is the total number of subjects and m is the number of predictor variables.

Stage B

Williams' modification of Hotelling's two predictor test is distributed approximately as F with 1 and N-3 degrees of freedom; it can be written as follows:

$$\frac{(r_{yx_i} - r_{yx_j})^2}{2\left[(1-R^2_{yx_ix_j})/N-3\right](1-r_{x_ix_j}) + \frac{(r_{yx_i} + r_{yx_j})^2 (1 - r_{x_ix_j})^3}{4 (N-1) (1 + r_{x_ix_j})}}$$

where r_{yx_i} and r_{yx_j} are the sample correlations between the dependent variable and the ith and jth predictor variables respectively.

N is the total number of subjects, $R^2_{yx_ix_j}$ is the sample coefficient of multiple determination between the dependent variable and the ith and jth predictor variables and $r_{x_ix_j}$ is the sample correlation between the ith and jth predictor variables.

Example

Several members of the faculty of the W.M.U. department of psychology were interested in the relationship between achievement in the required statistics course and performance on each of three predictor variables (G.R.E. - Verbal, G.R.E. - Quantitative and M.A.T.) sometimes used in the selection of graduate students. Measures on y (incorrect response on seven exams) and the three predictors x_1, x_2 and x_3 were obtained from each subject in a sample of 51 graduate students. The three predictor - criterion correlations were -.22, -.71, and -.01. Tests among these correlations follow:

Stage A

Step 1

$$\underline{R}_x = \begin{bmatrix} 1 & .311746 & .695068 \\ .311746 & 1 & .275878 \\ .695068 & .275878 & 1 \end{bmatrix}$$

Step 2

$$\underline{R}_x^{-1} \begin{bmatrix} 1.994826 & -.259082 & -1.315064 \\ -2.59082 & 1.116027 & -.127808 \\ -1.315064 & -.127808 & 1.949318 \end{bmatrix}$$

Step 3

$$\underline{r}_{yx} \begin{bmatrix} -.2247784 \\ -.7079091 \\ -.0113289 \end{bmatrix}$$

Step 4

$$R^2_{yx_1x_2x_3} = .56912087$$

Step 5

$$\sum\sum c_{ij} = 1.656263$$

Step 6

$$c_1 = .42068$$

$$c_2 = .729137$$

$$c_3 = .506446$$

Step 7

$$w_1 = \frac{c_1}{\sum\sum c_{ij}} = .25399935$$

$$w_2 = \frac{c_2}{\sum\sum c_{ij}} = .44023021$$

$$w_3 = \frac{c_3}{\sum\sum c_{ij}} = .30577632$$

Step 8

$$h = \begin{bmatrix} .25399935 & .44023021 & .30577632 \end{bmatrix} \begin{bmatrix} -.2247784 \\ -.7079091 \\ -.0113289 \end{bmatrix} = .37220065$$

Step 9

$$h^2\sum\sum c_{ij} = .13853332\ (1.656263) = .229448$$

Step 10

$$F = \frac{(.3396733)/2}{(.4308792)/47} = 18.53$$

Since the obtained value of F exceeds the critical value of F (which is 5.11 for alpha = .01 using 2 and 47 degrees of freedom) we reject the overall hypothesis

$$\rho_{yx_1} = \rho_{yx_2} = \rho_{yx_3}$$

and proceed to Stage B.

Stage B

Pairwise comparisons among the three predictor-criterion correlations are computed as follows:

$H_0\colon \rho_{yx_1} = \rho_{yx_2}$

$$F = \frac{\left[(-.2247784)-(-.7079091)\right]^2}{2\left[(1-.50115)/48\right](1-.3117462) + \left[\frac{\left[(-.2247784)+(-.7079091)\right]^2 (1-.3117462)^3}{4\,(51-1)(1.3117462)}\right]}$$

$$= \frac{.2334152}{.0143056}$$

$$= 15.17$$

$H_0\colon \rho_{yx_1} = \rho_{yx_3}$

$$F = \frac{\left[(-.2247784)-(-.0113289)\right]^2}{2\left[(1-.09115)/48\right](1-.6950676) + \left[\frac{\left[(-.2247784)+(-.0113289)\right]^2 (1-.6940676)^3}{4(51-1)(1+.6950676)}\right]}$$

$$= \frac{.0455606}{.0115519}$$

$$= 3.94$$

$H_0\colon \rho_{yx_2} = \rho_{yx_3}$

$$F = \frac{\left[(-.7079091)-(-.0113289)\right]^2}{2\left[(1-.53777)/48\right](1-.4758782) + \left[\frac{\left[(-.7079091)+(-.0113289)\right]^2 (1-.2758782)^3}{4\,(51-1)\,(1.2758782)}\right]}$$

$$= \frac{.4852239}{.0147158}$$

$$= 32.97$$

The three pairwise F tests yield values of 15.17, 3.94 and 32.97. These obtained values are compared with the critical value of F based on alpha = .01 with one and 48 degrees of freedom. Since the critical value is 7.25 we reject $H_0\colon \rho_{yx_1} = \rho_{yx_2}$ and $H_0\colon \rho_{yx_2} = \rho_{yx_3}$ and retain $H_0\colon \rho_{yx_1} = \rho_{yx_3}$.

Method II

The second method is proposed for the situation in which a relatively large number of predictor-criterion correlations is involved but the researcher has interest in making only a few planned comparisons. Unlike Method I, no preliminary overall test is run. Instead, each planned comparison is tested using Williams' modification of Hotelling's two predictor test (i.e., Stage B of Method I). The obtained F values are not, however, compared with the conventional points of the F distribution. Rather, the obtained F values are compared with the critical value of the Bonferroni F statistic associated with C (the number of planned comparisons) and one and N-3 degrees of freedom.

Example

If Method II is applied to the three correlations employed in the example previously described, we simply compare the obtained F values of 15.17, 3.94, and 32.97 with the tabled value of the Bonferroni F statistic (available in Huitema, In Press).

If all three pairwise contrasts are planned, we find the critical value is 9.59 for alpha (family) = .01. (If only two of the three possible pairwise contrasts had been planned we find the critical value is 8.69). Note that the decisions concerning the comparisons are the same (for these data) with Methods I and II.

Discussion

An important consideration in the application of Methods I and II is the experimentwise error rate. The probability of one or more false rejections in a study with m contrasts can be expected to be equal to or less than the nominal alpha with Method I if the Stage A test maintains the experimentwise error rate at the nominal alpha. Since the characteristics of the m predictor generalization of Hotelling's test (i.e., Stage A) have not been investigated, a clear statement of the error rate associated with this method is not currently possible. However, studies of the per comparison error rate associated with Hotelling's two predictor test suggest that the error rate is greater than alpha with certain correlation matrices. Hendrickson, Stanley and Hills (1970) and Hendrickson and Collins (1970) compared the empirical results of Hotelling's test with the results obtained using Williams' modification of Hotelling's test and Olkin's test. Both Williams' and Olkin's procedures were designed for the trivariate normal case rather than the fixed predictor case. The three procedures yielded practically the same results in twelve similar empirical studies. On the other hand, a very extensive Monte Carlo study carried out by Neill and Dunn (1975) led them to conclude that Hotelling's test is completely unsatisfactory because it fails to control Type I error near the nominal value with certain correlation matrices. The empirical Type I error rate was found to be as high as .92 when a nominal alpha value of .05 was employed. Williams' test maintained the empirical error rate very close to the nominal value regardless of the correlation matrix employed. Unfortunately, Neill and Dunn did not include Olkin's test in their study.

Since the m predictor generalization of Hotelling's test may also yield higher than nominal Type I errors with certain correlation matrices, the first stage of Method I may allow too many contrasts to reach Stage B. This will then lead to too many errors at Stage B to maintain the experimentwise error rate at the nominal value.

The experimentwise error rate associated with Method II is less than alpha because (a) Williams' test maintains the per comparison error rate at alpha and (b) the Bonferroni inequality makes it clear that the alpha associated with the whole collection of contrasts can not be greater than the sum of the individual alphas, i.e.,

$$\alpha_{\text{experimentwise}} \leq \sum_{i=1}^{m} \alpha_{\text{individual}} .$$

Since it is not yet clear whether or not Stage A of Method I sufficiently controls experimentwise error, the conservative approach is to employ Method II. Monte Carlo studies of the power and error rate associated with a large variety of correlation matrices are needed for both methods.

References

Hendrickson, G. F., and Collins, J. R. Note correcting the results in "Olkin's New Formula for the significance of r_{13} vs. r_{23} compared with Hotelling's method." American Educational Research Journal, 1970, 7, 639-641.

Hendrickson, G. F., Stanley, J. C., and Hills, J. R. Olkin's new formula for significance of r_{13} vs. r_{23} compared with Hotelling's method. American Educational Research Journal, 1970, 7, 189-195.

Hotelling, H. The selection of variates for use in prediction with some comments on the general problem of nuisance parameters. Annals of Mathematical Statistics, 1940, 11, 271-283.

Huitema, B. E. Three multiple comparison procedures for contrasts among correlation coefficients. Paper presented at the annual meeting of the American Statistical Association, St. Louis, 1974.

Huitema, B. E. Multiple comparison tests for contrasts among correlated correlation coefficients. Paper presented at the annual meeting of the American Statistical Association, Atlanta, 1975.

Huitema, B. E. The Analysis of Covariance and Alternatives. New York: Wiley, In press.

Marascuilo, L. A. Large-sample multiple comparisons. Psychological Bulletin, 1966, 65, 280-290.

Neill, J. J., and Dunn, D. J. Equality of dependent correlation coefficients. Biometrics, 1975, 31, 531-543.

Williams, E. J. The comparison of regression variables. Journal of the Royal Statistical Society - Series B, 1959, 21, 396-399.

A CODING APPROACH TO THE SENSITIVE QUESTION PROBLEM

Boris Iglewicz, Temple University

ABSTRACT

A coding approach, for handling sensitive questions in a group environment, is introduced. This is a variant of the direct questioning approach which still provides considerable protection to the respondents. An actual application of the coding approach is discussed.

1. INTRODUCTION

Many practical questionnaires contain questions which are of a sensitive nature. The respondent often either refuses to answer such questions or deliberately falsifies his answers. Such refusals to cooperate can cause large nonsampling errors which could lead to meaningless estimates of the desired parameters.

Warner [3] introduced a randomized response technique for increasing the rate of cooperation. The randomized response approach has recently received considerable attention. An excellent review paper on the subject has been written by Horvitz, et al. [1], where the properties of a number of randomized response techniques have been summarized. For the purpose of this note, it is sufficient to consider only Warner's technique, as the others have similar properties. Let A be a sensitive characteristic and Π_A be the proportion of the population possessing property A. Warner suggested that the interviewees be provided with a randomization device and be instructed to answer one of the following two questions with probability p and 1-p, based on the outcome of the randomization device, respectively:

1. I am an A.

2. I am not an A.

In this fashion, the interviewee's membership in A cannot be conclusively determined from the response. Warner's procedure leads to an unbiased maximum likelihood estimator $\hat{\Pi}_A$ of Π_A with

$$\mathrm{Var}(\hat{\Pi}_A) = (\Pi_A)(1-\Pi_A)/n + (p)(1-p)/[n(2p-1)^2], \quad p \neq \tfrac{1}{2}. \qquad (1)$$

Here n is the number of individuals interviewed. It can be seen that $\mathrm{Var}(\hat{\Pi}_A)$ consists of the variance based on direct questioning plus a second component due to randomization. The second term in equation (1) can be quite large when p is close to ½. The advantage of the randomized response technique occurs when the questions are truly sensitive. Randomization would increase the rate of cooperation for sensitive questions and, in this fashion, lead to a viable estimator. On the other hand, direct questioning would be preferred for nonsensitive questions.

Even though the models incorporating randomized response techniques are theoretically appealing, they have not been widely used in practice. Obviously, the randomized approach of indirect questioning is not as practical to use as to state. In this note, a variant of direct questioning is introduced which continues to provide a reasonable degree of protection and yet is easy to use. This technique is designed for groups of people, such as classes of students. An example is discussed in this note.

2. THE CODING APPROACH FOR DICHOTOMOUS QUESTIONS

The author had two back to back statistics I classes, consisting largely of undergraduate business students with rather weak quantitative backgrounds, making it almost irresistible not to experiment on them. The question was how to run a survey consisting of a number of sensitive questions? One constraint was that the survey should consume only a small amount of class time. In addition, it was felt that the students would be hesitant to respond, not only because they were fearful of the interviewer, but also because fellow students could glance at the responses. The latter problem cannot be wholly eliminated by the randomized response approach, as a fellow respondent may possibly notice the color of the marble chosen, from the randomization device, in addition to the final response. For these reasons a coded direct question approach became more appealing.

The survey was conducted on a typical class day. The questionnaire contained two groups of questions: the first group consisted of four rather nonsensitive demographic questions while the second group consisted of five sensitive questions. The four nonsensitive questions were also used as an instructional illustration of the coding procedure. The students were told that the author could decode their answers given a minute or so. They were also informed that the questions were to be distributed randomly and that the questionnaires were to be returned face down and then shuffled in order to avoid any possibility of identifying the interviewees. In other words, a deliberate effort was made to guard the privacy of the respondents.

The five sensitive questions used were:

5. I have used marijuana.

6. I have <u>never</u> used L.S.D.

7. I have used heroine.

8. I have <u>never</u> cheated on an inclass examination at Temple University.

9. I have professionally seen a psychiatrist or psychologist at least ten times.

The first four questions were nonsensitive. It was explained that questions 6 and 8 are asked in the negative, for further confusion to roaming eyes, and discussed how to reply to a double negative. The code used was for a person to lie on any two questions and then compute K + 2L and 2L + K, where L and K are the questions with numbers corresponding to the untruthful replies. One clearly now has two equations and two unknowns and can easily decode the responses. Furthermore, the respondents were asked to compute K + 3L and L + 3K in order to allow for a numerical error.

It is interesting to note that out of 66 students attending my class that day, 57 cooperated on the sensitive questions while nine did not. Failure to cooperate consisted of nonsensical answers, such as K + 2L = 100 and L + 2K = 100. It should further be noted that these students were rather young with 66.67% responding that they were less than 21 years old. The decoded percentages for the sensitive questions, and their standard errors are given in Table 1.

Table 1

Responses to Sensitive Questions

Question	Percentage Of Yes Responses	Standard Error
Used Marijuana	77.2	5.56
Used L.S.D.	14.0	4.60
Used Heroine	3.5	2.43
Cheated on Exam	63.2	6.39
Seen Psychiatrist	10.5	4.06

In this procedure the students were given the choice of the two questions to falsify. The fact that about half the answers were false was especially appealing to the respondents. If a survey were to contain only one or two sensitive questions, then a number of nonsensitive questions should be added in order to be able to use the coded procedure. Similarly, a large number of questions should be subdivided into smaller groups.

One argument against this coding procedure is that if a person were to refuse to answer one dichotomous question, all information for the other questions, in the group, would be lost. In this regard, it can be noticed that if questions are asked individually, then refusal to respond to a question would be incriminatory. For this reason, the person would either lie on this question or refuse to answer the entire questionnaire.

It should now be apparent that the direct questioning approach has some potential in sensitive question surveys. Comparisons between different survey techniques, such as the one performed by Smith, et al. [2], should include the direct questioning approach. In comparing randomized response techniques with direct questioning, the most ingenious direct questioning approaches should be used. The author hypothesized that in certain types of surveys direct questioning would prove superior, in terms of MSE, as compared to randomized response.

BIBLIOGRAPHY

[1] Horvitz, Daniel, G., Greenberg, Bernard, G., and Abernathy, James, R., "Recent Developments in Randomized Response Designs," in J.N. Srivastava, ed., A Survey of Statistical Design and Linear Models, New York: North-Holland Publishing Company, 1975, 271-285.

[2] Smith, L.L., Federer, W.T., and Raghavarao, D., "A Comparison of Three Techniques for Eliciting Truthful Answers to Sensitive Questions," Proceedings of the Social Statistics Section, American Statistical Association (1974).

[3] Warner, S.L., "Randomized Response. A Survey Technique for Eliminating Evasive Answer Bias," Journal of the American Statistical Association, 60 (1965), 63-69.

A MODEL OF LABOUR MIGRATION AND URBAN UNEMPLOYMENT IN AFRICA

Madugba I. Iro, University of Lagos
and
Lou. Fulkerson, Ahmadu Bello University, Zaria

1. Introduction

Rural to urban labour migration and the high rate of urban unemployment continue to create serious economic and social problems in the developmental processes of many African countries.

The relationships between migration and urban unemployment and the consequences of their interaction on labour mobility are complex and stand in need of clarification and quantitative analysis afforded by mathematical modelling, but few such models have been presented. Todaro (1969) has constructed such a model from the point of view of an economist. It is essentially a deterministic model. Our model, which is thoroughly probabilistic, is more sociological in its orientation, although we take heed of economic considerations. It predicts the pattern of movement of the labour force and can be used to compare the probable consequences of social and economic policies.

2. The Model

Suppose we consider the labour force in an African country as being divided into six states, as follows:

Rural	1. Unemployed (or underemployed) in the rural sector;
	2. Employed in agriculture;
	3. Employed in the rural sector, but not in agriculture;
Urban	4. Unemployed in the urban sector;
	5. Employed in the urban sector;
Rural or Urban	6. Departed from the labour force.

Further refinements can be envisaged, but it would not be fruitful to attempt a finer division than can actually be observed. We assume that it is possible to count the number of people who move from each state to each other state and thereby to estimate the probability of making such a move. This information we summarise in a matrix of transition probabilities, $P=(p_{ij})$, $i,j = 1,\dots,6$.

The model as formulated so far is, of course, a finite Markov-chain. But this structure is too coarse. It does not take into account variation of time of stay in each state, which is an important factor in a realistic analysis of migration. Thus, it is a well-known fact of social life that the longer one has been in a particular state the less likely one is to move to another state (the so-called Axiom of Cumulative Inertia) (McGinnis, 1968).

We incorporate these realities into our model by assuming that for every pair of states (i.j) there exists a distribution $F_{ij}(t) = P_r$ going from i to j on or before time t, given that we are going to make that move. For such a process one can calculate numerous quantities of interest, such as the means and the variances of first passage times (the times to go from state i to state j, for example the time for the unemployed country boy to get a job in the city).

If the "embedded Markov-chain" of our process is ergodic, we can calculate the limiting probabilities of each state, the proportion of time which the process spends in each state, in the long run.

For those calculations, we need to know only first and second moments of $F_{ij}(t)$, which we can estimate by collecting data on how many people moved from i to j and how long they waited before doing so.

3. A further assumption on the $F_{ij}(t)$

We have assumed, at least tentatively, that the $F_{ij}(t)$ are gamma distributions, with density

$$f_{\alpha}\,v(x) = \frac{1}{\Gamma(v)} \alpha^{v} x^{v-1} e^{-\alpha x}, \quad x > 0.$$

For $0 < \alpha < 1$, the gamme distribution is a "decreasing failure rate" distribution, consistent with the "axiom of cumulative inertia." For V=1, it reduces to an exponential distribution. Since some of our transitions will no doubt resemble poisson processes, this generality of the distribution form is appropriate.

We can estimate from a sample the parameters of the $F_{ij}(t)$, evaluate them by numerical integration, and so evaluate the <u>renewal quantities</u>.

$$W(t) = (I - P \times F(t))^{-1} - I$$

(The "x" denotes element-by-element matrix multiplication).

$W_{ij}(t)$ is the mean number of moves into j in (o,t], given that we are in state i at time 0, so that knowledge of W(t) is equivalent to knowledge of how many people are in each state at time t if we know how many people were in each state at time 0.

4. Causal Structure

The semi-Markov model which we have presented is <u>descriptive</u>, although in a quantitative fashion. Given data, it will give us information about what is likely to happen.

We can, as has been suggested by Ginsberg (1972), incorporate causal structure within the semi-Markov framework by expressing some of the parameters of the semi-Markov process as functions of observable "exogenous" variables and seeking to estimate the parameters of these functions, which we shall call "causal functions."

Considering the transition matrix, we can think of the P_{ijs} as being functions of known form of exogenous variables $x_1, \dots, x_q$ determined by parameters $B_1, \dots, B_r$ which we must estimate.

We might have, for example,

x_1 = private investment in the rural sector;
x_2 = public investment in the rural sector;
x_3 = private investment in the urban sector;
x_4 = public investment in the urban sector;
x_5 = investment in education in rural areas;
x_6 = investment in education in urban areas.

We shall take the same attitude toward the moments of F (t), or specifically, for reasons which will come out presently, the unconditional mean waiting times.

$$m_i = \Sigma P_{ij} m_{ij},$$

where $M = (m_{ij})$ is the matrix of means of F (t).

In our initial investigations we shall assume that the causal functions are of the form

$$P_{ij} = 1 - \exp\left(-\sum_k B_{ijk} x_k\right)$$

If we have estimated transition matrices $P^{(\ell)}, \ell = 1, \ldots, N$, where the values of ℓ might perhaps correspond to political or geographical divisions, and we have corresponding values of $X^{(\ell)} = (x_1^{(\ell)}, \ldots, x_r^{(\ell)})$ we shall have N equations in r unknowns, an over-determined system which we can solve in the least-squares sense for the B's.

Ginsberg (1972) shows how one can get maximum likelihood estimates of various types of causal functions. Unfortunately, his specific proposals are not directly applicable to our case. He thinks of people making choices of where to move to (or not to move at all). In Nigeria for example, his states might be the nineteen states of the Federation. Ginsberg regards the P_{ij} as increasing functions of the attractiveness of destination j and decreasing functions of a distance, geographical, financial, or social, (d_{ij}). Given our definition of states, the notion of distance is not relevant, and our people have limited choice. Few of them would choose to become unemployed, and fewer still would choose to quit the labour force by illness, death, or going to jail.

5. The Model as an Instrument of Policy

We should like social analysis to be helpful in the rational formulation of social policy. If we can express at least some of the parameters of the semi-Markov process as functions of variables which are accessible to policy, we can move in this direction by using the techniques of Markov-renewal programming.

In this model, or from our point of view, submodel, if we make a move from i to j in time Δt, we receive a reward, positive, negative, or zero, (which in general is conditional on i, j, and Δt). By an iterative scheme of the dynamic programming type, we can find a "policy", i.e. a set of parameters which will maximise either average reward in the long run, if the embedded Markov-chain is ergodic, or expected total reward in transient states, if there are absorbing states.

6. Some Preliminary Results of the Markov-Renewal Programming Model

We have coded the Markov-renewal programming for a computer, in APL, and tried it, using the data shown below. The transition probabilities and mean waiting times in Tables I and II are from a study done by one of our students, Mrs. Ezim Okeke. She considered only five states,

1. Employed in a rural area
2. Unemployed in a rural area
3. Employed in an urban area
4. Unemployed in an urban area
5. Departed from the labour force

She regarded state 5 as being an absorbing state (dead or retired).

The transition probabilities might have been affected by factors such as seasonal changes in employment – occasioning movement away from farming to trading.

Table I: Matrix of Transition Probabilities (P_{ij})

$$P = \begin{bmatrix} 0.79 & 0.01 & 0.10 & 0.05 & 0.05 \\ 0.03 & 0.87 & 0.00 & 0.08 & 0.02 \\ 0.10 & 0.18 & 0.65 & 0.03 & 0.04 \\ 0.15 & 0.10 & 0.16 & 0.55 & 0.04 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Table II: Matrix of Waiting Times

$$M = \begin{bmatrix} 0.90 & 0 & 0.05 & 0.05 & 1.00 \\ 0.10 & 0.65 & 0.07 & 0.18 & 0.85 \\ 0.45 & 0.05 & 0.45 & 0.05 & 0.60 \\ 0.05 & 0.48 & 0.20 & 0.27 & 0.65 \\ 0 & 0 & 0 & 0 & 100.00 \end{bmatrix}$$

The only reason that observational data are needed at all for the Markov-renewal program algorithm is that in practice we want to fix some of the transition probabilities, saying that they represent phenomena which are beyond the reach of public policy.

The matrix displayed in Table III below tells the computer program which probabilities are to be fixed: these are the non-negative entries which, obviously, are identical to the corresponding elements of the transition probability matrix of Table I. The negative numbers in Table IV tell the program where to calculate new probabilities in these positions as elements of an optimal policy.

In this case, in which state 5 is absorbing, the optimal policy is a matrix of transition probabilities and a vector of unconditional mean waiting times which will maximise the total reward earned in passage through transient states. The reward structure is defined by a matrix, an example of which is displayed in Table IV.

Table III: Computed Policy

0.95	0	0	0	0.05	0.855
0.98	0	0	0	0.02	0.098
0.10	0.18	0.68	0	0.04	0.306
0.15	0.10	0.71	0	0.04	0.142
0	0	0	0	1	1

Table IV: Reward Matrix

80	-20	-10	-800	0
30	-20	10	-600	0
20	-10	20	-400	0
160	20	30	-800	0
0	0	0	0	0

The reward matrix expresses our value judgments as to the relative merits of the possible transitions. The matrix of this example says that it is good to stay on the farm or to return to it; bad to move to the city; worse to move to unemployment or to lose one's job, good to get a job. We are indifferent to departure from the labour force.

CONCLUSION

The semi-Markov model enjoys the virtues of flexibility and adaptability. The model does not care what or how many states it has; neither does it care about our choices of the functions by which we relate the parameters of the semi-Markov process to other observable variables or the means by which/estimate the parameters of these functions: we can choose functional forms and estimation procedures as the data and our intuitions about the pertinent social processes direct. Having established such dependencies to our satisfaction, we can use the Markov-renewal programming algorithms to explore the consequences of resource allocation decisions, thus making, at least potentially, practical use of the model.

Human beings are free agents, and their behaviour is stochastic. The semi-Markov model reflects this reality. Within its probabilistic framework we can incorporate modelling of causal structure, expressing some aspects of behaviour as functions of observable, "exogenous" variables while leaving others to direct statistical estimation.

We cannot expect that these techniques will have the neat usefulness that linear programming has to the operation of a refinery; nonetheless, any means by which quantitative estimates of the probable effects of policy may be had should not be scorned. The makers of public policy patently need all the help they can get. A famous example of how a well-meant policy can miscarry is the attempt of the Kenyan Government to reduce urban unemployment problem by subsidising wages, on agreement with employers to increase their labour force. Promptly many, many people rushed to urban areas, and the urban unemployment problem became worse than ever. It is conceivable that the construction of urban housing by the Nigerian Government has had a similar side effect: more shelter available leads to more relatives coming to town.

These plans were made by reasonable men doing their best to choose the least evil, but had they had more means of extrapolation they might have done better.

REFERENCES

1. Barlow, R.E. & Proschan, R.: Mathematical Theory of Reliability, New York, Wiley, 1965.

2. Businger, P.A. & Golub, G.H.: Linear least squares solutions by Householder transformations, Numerische Mathematik, 1965, Vol. 7, 269-276.

3. Ginsberg, Ralph B.: Semi-Markov processes and mobility, Journal of Mathematical Sociology, 1971, Vol. 1, 233-262.

4. Ginsberg, Ralph B.: Critique of Probabilistic models: Application of the semi-Markov model to migration, Journal of Mathematical Sociology, 1972, Vol. 2, 63-82.

5. Ginsberg, Ralph B.: Incorporating causal structure and exogenous information with probabilistic models: with special reference to choice, gravity, migration, and Markov-chains, Journal of Mathematical Sociology, 1972, Vol. 2, 83-103.

6. Howard, R.A.: Semi-Markovian decision processes, Revue et Bulletin Institut International de statistique, 1963, Vol. 40, 625-652.

7. Hunter, G.: Rural-urban migration in Tropical Africa, Commonwealth Lecture, 1974.

8. Jewell, W.S.: Markov-renewal programming, Operations Research, 1963, Vol. 2, 938-971.

9. McGinnis, R.: A Stochastic Model of Social Mobility, American Sociological Review, 1968, Vol. 33, 712-722.

10. Pyke, R.: Markov-renewal processes: Definitions and preliminary properties; Markov-renewal processes with finitely many states, Annals of Mathematical Statistics, 1961, Vol. 32, 231-259.

11. Schweitzer, Paul J.: Perturbation Theory and undiscounted Markov-renewal programming, Operations Research, 1969, Vol. 17, 716-729.

12. Schweitzer, Paul J.: Private communication.

13. Todaro, M.: A model of labour migration and urban unemployment in less developed countries, The American Economic Review, 1969, Vol. 59, No. 1, 138-148.

14. Todaro, M. & Harris, J.R.: Migration unemployment and development: A two-sector analysis, American Economic Review, 1970, Vol. 60, No. 1, 126-142.

15. Todaro, M.: Income expectations, rural-urban migration, and employment in Africa, International Labour Review, 1971, Vol. 104, No. 3.

INTERNATIONAL PERINATAL MORTALITY: COMPARABILITY OF DATA

Robert A. Israel and Anders S. Lunde*
National Center for Health Statistics

Accurate and comprehensive annual infant mortality rates are computed for only a minority of the world's population, comprised of most of the countries of Europe and North America, several countries in Latin America, Africa, and Asia, and Australia and New Zealand in Oceania. Infant mortality measures of varying degrees of unreliability exist for many other countries, and for the rest of the world such data are nonexistent. Similarly, perinatal mortality a measure resulting from the combination of appropriate subsets of both fetal and infant mortality, is available for study on a reliable basis for only a relatively few countries and for a small proportion of the world population. Nevertheless, in many of the more developed countries, perinatal deaths now exceed in number the deaths in the next 30 to 40 years of life, while in less developed countries the toll of such early mortality is usually significantly greater. However, even for those countries where perinatal mortality data are available, questions arise concerning terminology and definition, quality and completeness of data, and analytical techniques and methodology, all of which affect to a greater or lesser degree the value of the information for use within countries, for international comparisons, or for analysis of trends through time. Nonetheless, the continuing efforts of the United Nations and the World Health Organization to encourage nations to improve their vital statistics, and the growing interest of the nations themselves in developing statistics for national planning, have resulted in an increase in countries producing fetal death statistics, neonatal statistics, and subsequently, perinatal statistics for health status analysis.

Another reason for the recent surge of interest in perinatal statistics is the significant decline in the perinatal mortality rates during the last decade in a number of countries, particularly in Scandinavia. This represents a change from the previous decade in which a decrease in the rate of decline over previous periods had been observed. In the period 1955-64, for example, the decrease in the perinatal mortality rate for the United States was 5.7 percent, while in the decade 1965-74 the rate dropped by 31.1 percent. Similar declines were noted for Canada and the countries of Northwestern Europe. Of interest by way of contrast, the perinatal mortality rate in Hungary, which showed a similar pattern to that of the United States of a slowdown in the rate of decrease during the period 1955-1964, has not experienced a resumption of the accelerated decline. These same trends noted in perinatal mortality were first noted in the patterns of infant mortality for these countries during the same period, after leveling off in the 1950's and 1960's, the U.S. infant mortality rates unexpectedly fell from 24.7 in 1965 to 16.7 in 1974, a decline of some 32 percent. 1/

It may be said that while for most of the world the general infant mortality rate remains a primary social indicator, in developed countries the perinatal mortality rate is replacing it. A number of countries which hitherto have been unable to produce adequate data to calculate perinatal mortality rates are now attempting to do so and the development of perinatal mortality statistics is recognized as a goal for many of them. That the perinatal period is a highly significant time in the study of early mortality is demonstrated by the fact that in Norway, for example, where both civil and medical registration of births and deaths is quite complete and uniform throughout the country "more than 80 percent of deaths before and after birth and up to one year of life, compiled in official statistics now occur in the perinatal period."2/

The purpose of this paper is to examine the trends in perinatal mortality statistics in selected countries, to review the problems of measurement and comparability, and to suggest further areas for research and analysis.

Components of Perinatal Mortality

The components of perinatal mortality as used in this paper are late fetal deaths (deaths of 28 weeks or more gestation) and semanatal deaths (deaths occurring in the first six days of life). In recognition of the fact that mortality during late gestation periods is closely allied to mortality in the early neonatal period, especially as regards cause of death, the WHO in 1954 supported the establishment of a perinatal mortality rate. Since that time WHO expert groups have been at work developing standards and assessing progress in international perinatal mortality measurement. There are two kinds of perinatal mortality measures in use: perinatal mortality rates and perinatal mortality ratios. The perinatal mortality rate is defined as the number of deaths under one week of age plus late fetal deaths (perinatal deaths) per 1,000 live births and late fetal deaths. The perinatal death ratio is the number of perinatal deaths per 1,000 live births. In other words, the rate includes late fetal deaths in both numerator and denominator. The inclusion of counts of both live births and late fetal deaths in the denominator serves to more closely approximate the population at risk of dying. The most appropriate denominator, were such a statistic available, would be a count of

*Paper contributed at the annual meeting of the American Statistical Association of America, Boston, Massachusetts, August 24, 1976. Mr. Israel is Deputy Director, National Center for Health Statistics, and Dr. Lunde is Director, Office of International Statistics, NCHS. The opinions expressed are those of the authors and do not necessarily reflect the views of the National Center for Health Statistics and the Department of Health, Education, and Welfare.

total pregnancies. The actual numerical difference between rates and ratios in this case is slight, at least among countries where registration is fairly well complete. For example, in 1974 the U.S. perinatal rate was 18.9 and the ratio was 19.1, a difference of 0.2 points or 1 percent. The measure referred to most frequently in this paper will be the perinatal mortality rate on the basis that the denominator comes closer to estimating the true population at risk. The UN Demographic Yearbook publishes international perinatal mortality data in terms of ratios; it is considered that having only births in the denominator lends a certain amount of stability to the measure.

Trends of the Perinatal Mortality Rate

Perinatal mortality declined in the countries of low infant mortality throughout the last 50 years. Chase noted that the rate of decline in 1950-65 had been slower than in previous periods.3/ Between 1955 and 1964, of these countries, Denmark showed a rapid decline of 31 percent and England and Wales, Finland, and Sweden declined by around 20 percent. In Hungary the decline was only 12 percent and in the United States it was practically stationary during that period.

In the next decade, 1965-1974, a significant change was noticeable. There was an increase in the rate of decline and Denmark's perinatal mortality rate fell by almost 40 percent and Finland's by 37 percent. In Sweden, the Netherlands, and the United States the rates dropped about a third. The region of Northwest Europe was not entirely consistent, as Norway declined by 29 percent and England and Wales by around 22 percent. However, the rate for Hungary during this period was almost stationary. When one examines the component rates of fetal mortality and semanatal mortality, one can see that these rates declined accordingly. Lower late fetal deaths had a greater impact than semanatal mortality levels in Sweden and England and Wales, while semanatal declines had the greatest impact in Canada, Finland, the Netherlands, and Norway.

Problems of Measurement

Difficulties in comparing statistics from one country to another or even within countries often stem from differences in definitions in terminology, and in the practice of applying the definitions and terms to specific cases. In recognition of this problem WHO has promulgated standard definitions of live birth and fetal death.4/ However, even with internationally recommended definitions for the most important terms, considerable international variations still exist. These variations, along with a discussion of the problems of application of national definitions within a country, are presented in the Handbook of Vital Statistics Methods.5/ But even when definitions appear to be the same or essentially similar, interpretation and practice in their application may affect the data to an unknown degree. Despite their socio-cultural similarities, the Scandinavian countries have shown variation in various mortality rates and have questioned the degree to which such differentials may be affected by differences in definitions and procedures. Bolander and Lettenstrom indicate that through joint efforts promulgated by the Nordic Medico-Statistics Committee (NORMESCO) in Norway, Sweden, Denmark, and Finland, adherence to common standards is being achieved and a better basis for comparability is being established.6/

In the measurement of perinatal mortality, all the limitations of fetal death statistics, infant death statistics, and live birth statistics must be considered in evaluating international comparability. One major problem, related to the imperfect registration of live-born infants, is eliminated in perinatal statistics: the question of whether the product is a live birth or a fetal death. Still, exclusion from live birth statistics in some countries of those live born who die shortly after birth and are registered as stillbirths can have the effect of increasing the ratios by decreasing the number of live births in the denominator. Unreliable birth figures in some countries have inflated the ratio. Live births tabulated by date of registration rather than date of occurrence can have a similar effect. In some parts of Africa, Asia, and Latin America live birth statistics by date of registration (usually extensively delayed) produce birth rates of great magnitude. Incompleteness of current registration is hidden by the inclusion of events that occurred in prior years.7/

Fetal death statistics are probably the most unreliable of all vital statistics; because of the impossibility of determining the completeness of early fetal deaths (less than 20 weeks gestation), and uncertainties with intermediate fetal death (20 but less than 28 weeks gestation) only data on late fetal deaths are available for many reporting countries. In the UN Demographic Yearbook for 1974 data were available from only 45 countries or areas and the quality of much of the data may be questioned. There are also many differences in interpreting fetal death which make comparison difficult, despite WHO efforts to promote a standard definition based on evidence of life. Viability, for example, is defined in some countries in terms of length of gestation period ranging from 3 months to 7 months. Some countries have an additional requirement that fetuses be of a minimum length, from 30 to 35 centimeters. Others specify that the product show "definite" signs of life or some other nonspecific term.

Furthermore, there is now a growing dissatisfaction with the currently recommended definitions. For example, there is concern that the signs of life listed in the WHO definition of live births call for the inclusion as live births of very early, and patently non-viable fetuses, who may show one or more of the definite signs of life after such procedures as therapeutic termination of pregnancy in the early weeks of gestation. In addition, the criteria themselves are clearly subjective and open to various interpretations. Objections have also been voiced regarding the definition of fetal death which is supposed to be the converse of the definition of live birth and therefore should include all terminations of pregnancy other than the delivery of a live born infant. However, the current definition refers to the absence of signs of life in a fetus at the time of its separation from the mother. In many cases, the fetus is not available for examination at that time, if at all; and again, the criteria are subjective.

Another problem is related to differences in the registration requirements for fetal death. All of the States of the United States require registration of all other fetal deaths, or those of 20 weeks or more gestation. Most countries confine registration to fetuses of 28 weeks or more gestation. The general effect of this latter practice would be to record fewer fetal deaths because of the tendency to underestimate gestational age slightly over 28 weeks to avoid registration of fetal deaths. When, in the United States, registration requirements were lowered from 28 to 20 weeks and, later, in some ten states lowered from 20 weeks to all fetal deaths regardless of gestational age, the reporting at 28 weeks and over improved.

Underregistration of fetal deaths is a significant problem, even in developed countries. Registration completeness may account for the principal differences among the comparatively low perinatal rates in advanced countries.

Another problem in most countries is determination of gestational age. Where fetal deaths are reported there is frequently a large proportion of unknown gestation. Where these should be allocated is a question; usually these are allocated in statistical tables under later fetal deaths, which have an effect on the fetal death rates and ratios.

As regards semanatal deaths, in some countries death statistics are not available in terms of the first week of life; mortality is classified in terms of under 10 days or under 14 days of life. In some countries infants dying in the first week of life are excluded if they have not been registered while alive.

Needed Research

While it is generally true that the importance of measures of fetal, infant, and perinatal mortality is well recognized, the actual scope of these data vary considerably from country to country. When available at all, they range from the most basic establishment of the fact of death to sophisticated data systems designed to elicit information on the biologic, socio-economic, cultural and geographic determinants of early mortality. The scope of the data may include only those deaths occurring after birth (i.e. infant and childhood mortality) or may include fetal losses as well. If fetal deaths are included in a data collection system, they may be limited to those events occurring at 28 or more weeks of gestation. In this connection, a joint United Nations/WHO meeting8/ noted that probably at least 15 percent of all human individuals die before they have reached the twenty-eighth week of prenatal life. It also noted that the study of spontaneous fetal death at ages earlier than 28 weeks would be valuable since appreciable numbers of such early fetuses, when delivered, are now surviving. Attention is also being focused on early fetal losses due to legal abortion. Some countries, Hungary, for example, have adopted very liberal abortion laws but little is yet known about the effects of these policies on such measures as maternal mortality, perinatal, infant and childhood mortality, and on the demographic characteristics of the surviving population.

Similar to variations in the scope of the collected data are variations in the objectives for their analysis and use. While the ultimate goal of a statistical system dealing with mortality in early life must be the reduction of this mortality to its minimum, the attainment of such a goal is inevitably dependent upon the accomplishment of numerous intermediate objectives. Where measures of perinatal mortality are unreliable or unavailable, the first objective must be the development of a basic series of reliable, internationally comparable statistics. A second level of priority should be the maintenance of such measures on a periodic basis in order to assess changes in the levels of the measurements. Further priorities should deal with the collection of important additional variables that would allow further analysis and interpretation of the data. The completeness, accuracy and availability of such information will greatly assist in the evaluation of existing programs and future projects which are designed to meet the ultimate objective.

The sources from which the necessary information is derived are also somewhat varied, but, for the most part, countries are dependent upon their vital statistics systems to provide mortality and natality data. The vital statistics are, in turn, almost invariably derived from a system of civil registration, although sample survey methods and sample registration areas are employed in some countries as a substitute for complete registration. Surveys are also used in some countries in order to obtain additional data not available through the registration system. In some cases, a population enumeration or census has been used to collect information on vital events, but this method has not been generally satisfactory, primarily because it is retrospective and relies on the recall of the respondent and because such censuses are rarely conducted more frequently than once every five or 10 years. A few countries have established continuous population registers for the purpose of recording many types of civil events occurring to individuals. Such a continuous population register can provide the same kind of data as are available in the more limited vital statistics registration system, but usually provides additional opportunities for compiling and analyzing data through the mechanism of record linkage. Record linkage is, of course, a possibility with other data collection systems, but is usually much more difficult and costly. Other sources of data include hospital, health insurance, and physician records, but data obtained from studies based on these latter kinds of documents are usually too limited and atypical to be of general applicability. In reviewing the registration problem in Latin America and elsewhere, after noting the large proportion of deaths during the first day of life that were unregistered, Puffer and Serrano concluded that until hospital procedures are improved and standard definitions followed, comparability of perinatal death statistics will remain in doubt for most of the world.9/ Special studies, however, can provide data for perinatal mortality statistics when such data may be incomplete or lacking. For example, Laurenti was able to analyze perinatal mortality in Sao Paulo, Brazil, by combining all death certificates of late fetal deaths with a sample of infants under one year of age obtained from physician and hospital records. By far the most common sources of data on perinatal mortality are the officially registered documents of birth, death, and fetal death. However, most of the comments which follow regarding these documents are also applicable to other sources of data as well. One aspect that must be raised is that of the items or topics which are included on the records. In order to satisfy both national and international needs, efforts should be made in all countries to include certain basic topics irrespective of any additional items that may be desirable and practical in each country. Guidelines for these topics as well as a suggested tabulation program have recently been put forward.11/ Further consideration of the sources of data must include mention of the persons who supply the information for the Official Records. There are numerous possibilities depending on the practices, procedures and legal requirements of each country and the circumstances surrounding each specific case. The knowledge and qualifications of these informants have a pronounced effect on the final statistics to be derived from the system.

In addition to the kinds of data on early mortality that have been discussed, there is another area that remains almost completely uncovered by statistics, namely early fetal deaths or abortions. The problems of recognizing early pregnancy wastage and of encouraging reporting when such losses are recognized are so great that it is unlikely that adequate statistics can be obtained in any country in the foreseeable future. However, in some countries relaxation of the abortion laws makes the number of legally induced abortions a significant and measurable statistic. While this number has little value in assessing the total of all early pregnancy wastage, both induced and spontaneous, it is of value in the evaluation of the effects of medical intervention on the birth rate and on perinatal morbidity and mortality. Data for such statistics are likely to be the product of a specially designed reporting system. However, improved or new techniques 12/ in the survey field are now raising the possibility of collecting sensitive information about such topics as abortion as a part of population surveys.

The existence of a data collection system, however, does not, in itself, ensure accurate, reliable statistics. In order to understand and use data properly, there must be an appraisal of the quality of the collected information. Such an evaluation should cover a number of aspects such as:

1. What pertinent terms and definitions are in use?

2. Have these terms and definitions been uniformly applied throughout the time period for which the data have been collected?

3. Have these terms and definitions been uniformly understood and used in all geographic areas from which the data are collected?

4. How complete are the counts of the vital events of interest?

5. If the vital statistics system is not based on the registration of all events, but relies on sample surveys or other procedures to estimate the counts of vital events, how reliable are the estimates?

6. What are the sources and how accurate are the population bases used in the calculation of the mortality rates?

7. What is the proportion of completeness for each item or topic of interest on the data collection record?

8. What is the accuracy of each of these items or topics?

Answers to some of these questions are more readily ascertained than are others. For example, determination of the official definitions and whether they have been modified during a given period is an easier task by far than is the determination of how these terms and definitions are actually used in practice. The evaluation of the completeness of coverage of vital events is difficult at best, but might be done through routine independent sources, field investigations, or through various other demographic analytic techniques. Estimates of vital events should be accompanied by a measure of sampling error or by some other indicator of their accuracy. The accuracy of individual items or topics can be investigated by several techniques, including an analysis of the proportion of records where the response to a particular item is incomplete, inconsistent, or unknown. Independent estimates of certain items might be obtained through special surveys or from census data. Various consistency checks with these latter sources of data are sometimes possible and, if so, highly desirable. Often it is possible to assess the approximate degree of accuracy of an item merely on the basis of the magnitude of the statistic compared with level of the same statistic observed in other countries with similar characteristics at approximately the same point in time. It should be noted that evaluation of data is of fundamental importance, but as yet it is frequently overlooked. On the other hand, data collection systems of the type we are concerned with usually have many shortcomings, which, if recognized, can be tolerated or compensated for. Perfect, or even near-perfect statistics constitute an ideal which will not be realized. In a similar connection, Greenwood remarked: "The scientific purist, who will wait for medical statistics until they are nosologically exact, is no wiser than Horace's rustic waiting for the river to flow away."13/

Concluding Statement

That perinatal mortality is an important index of not only early mortality but of general health and wellbeing is widely accepted. However, the measure suffers from limited availability and problems of comparability around the world.

This paper presented some illustrative perinatal mortality statistics from a few countries with traditionally low mortality rates and generally reliable statistics. Even among these countries, it is difficult to ascertain how much of the difference in rates is due to real differences and how much is due to comparability issues. However, through the efforts of WHO and civil registration authorities, the medical profession, and other concerned persons in many countries, more reliable fetal and infant mortality data are becoming available. As civil registration systems improve and more accurate and complete data are tabulated it will become possible to extend production of the details of peri perinatal mortality statistics. The task in developing countries will be to improve registration completeness including those variables essential to basic fetal death, live birth, and infant death analysis. The task for developed countries will be to standardize registration terminology and practices to improve cause of death statistics, and to link fetal, births, and semanatal records with other socio-economic data for more comprehensive analysis.

REFERENCES

1. Lunde, A. S., Assessing international data on infant mortality. Unpublished paper presented at the annual meeting of the Population Association of America, Montreal, Canada, April 28-May 1, 1976.

2. Dødeligheten Omkring Fodselen og i Første Levear 1969-1972 (Fetal and Infant Mortality 1969-1972). Oslo: Central Bureau of Statistics of Norway, Statistiske Analyser Nr. 15, 1975, p. 60.

3. Chase, Helen, International Comparisons of Perinatal and Infant Mortality: The United States and Six West European Countries. Washington, D.C. NCHS Vital and Health Statistics, Series 3, Number 6, March 1967, pp. 34-35.

4. Manual of the International Classification of Diseases, Injuries, and Causes of Death, 1965 Revision, Vol. 1, Geneva, WHO, 1967, p. 469.

5. Handbook of Vital Statistics Methods New York: United Nations, Statistical Office, Department of Economic and Social Affairs, April 1955.

6. Bolander, A.M. and G. Lettenstrom, Comparative study of prenatal, infant and childhood mortality in four Nordic countries. International Population Conference, Liege 1973, Volume 3, pp. 143-156.

7. UN Demographic Yearbook 1974. New York: United Nations, Statistical Office, Department of Economic and Social Affairs, pp. 40-41, and footnotes to Table 19, "Perinatal Deaths and Perinatal Death Ratios, by Urban/Rural Residence 1965-73."

8. Programmes of analysis of mortality trends and levels. WHO Technical Report Series 440, Geneva, 1970.

9. Puffer, Ruth R. and C.V. Serrano. Patterns of Mortality in Childhood. Washington, D.C.: Pan-American Health Organization, 1973,p. 35.

10. Laurenti, R., et al. Mortalidade perinatal em Sao Paulo, Brasil. Rev. Saude publica, Sao Paulo. 9:115-24, 1975.

11. Principles and recommendations for a vital statistics system. Background paper submitted to the Sixteenth Session of the Statistical Commission. United Nations: UN.E/CN.3/411, 1969.

12. Warner, S.L. Randomized Response: A survey technique for eliminating evasive answer bias. JASA, Vol. 60, p. 309, March 1965.

13. Greenwood, M. Medical Statistics from Graunt to Farr. Cambridge: 1948, p. 28.

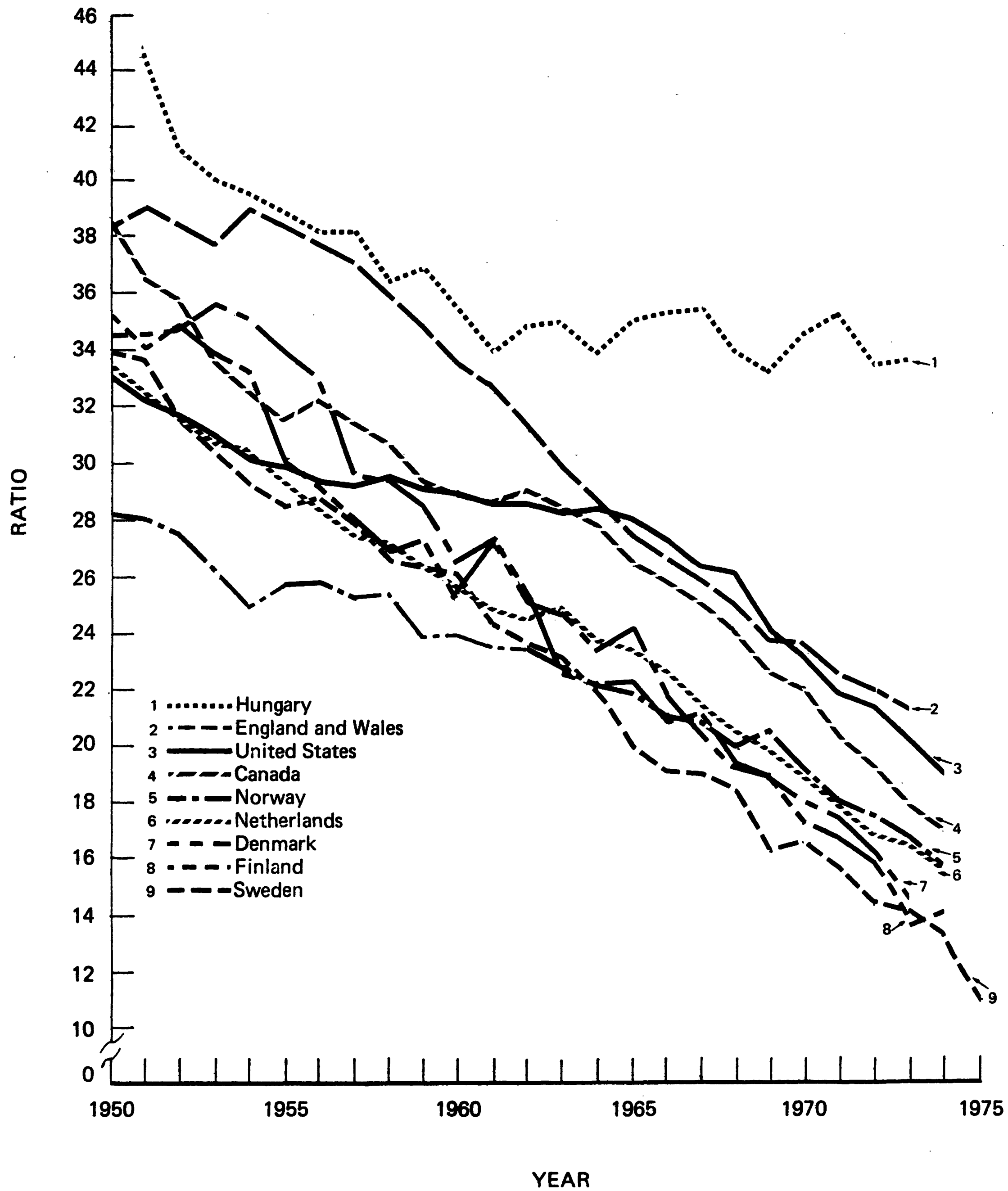
TABLE I. PERINATAL MORTALITY RATIOS, UNITED STATES AND SELECTED COUNTRIES, 1950-1975

Year	Country								
	USA	Canada	Denmark	England & Wales	Finland	Hungary	Netherlands	Norway	Sweden
1975	-	-	-	-	-	-	-	-	11.1e
1974	19.1	16.9	-	-	14.0e	-	15.5	15.6	13.3
1973	20.2	17.7	14.6	21.3	13.7	33.6	16.4	16.7	14.1
1972	21.4	19.2	16.2	22.0	15.8	33.4	16.7	17.6	14.4
1971	21.9	20.3	17.5	22.5	16.7	35.2	17.8	17.9	15.7
1970	23.2	22.0	18.0	23.8	17.2	34.5	18.8	19.3	16.5
1969	24.2	22.5	18.9	23.7	18.9	33.2	19.8	20.5	16.3
1968	26.1	24.0	19.1	25.1	19.3	34.0	20.4	19.9	18.4
1967	26.5	25.0	-	25.8	21.1	35.4	21.4	20.8	18.9
1966	27.2	25.8	21.8	26.7	20.8	35.3	22.7	21.1	19.0
1965	28.0	26.3	24.2	27.3	22.3	35.0	23.4	21.9	19.9
1964	28.4	27.7	23.4	28.6	22.2	33.8	23.7	22.1	21.9
1963	28.2	28.4	24.6	29.8	22.5	35.0	24.9	22.8	23.1
1962	28.5	29.0	25.2	31.4	25.3	34.8	24.4	24.0	23.6
1961	28.6	28.5	27.3	32.7	27.1	34.0	24.8	23.5	24.2
1960	28.9	28.8	26.5	33.5	25.3	35.5	25.6	24.0	26.2
1959	29.1	29.3	28.6	34.8	27.3	37.0	26.3	23.9	26.3
1958	29.6	30.6	29.4	35.8	26.9	36.3	27.1	25.4	26.5
1957	29.2	31.4	29.5	37.0	27.8	38.2	27.4	25.1	27.8
1956	29.3	32.2	33.0	37.6	29.2	38.1	28.3	25.9	28.8
1955	30.0	31.5	33.9	38.3	30.0	38.7	29.3	25.9	28.4
1954	30.2	32.5	35.1	39.0	33.3	39.5	30.4	24.7	29.2
1953	31.0	33.6	35.6	37.7	33.8	40.0	30.6	26.2	30.3
1952	31.6	35.8	34.6	38.3	34.7	41.0	31.5	27.5	31.5
1951	32.2	36.4	34.5	39.0	34.0	44.3	32.4	28.1	33.6
1950	33.0	38.6	34.5	38.3	35.1	-	33.4	28.2	33.8

e Estimated
- Data not available

Sources: 1950-74 U.S.A.: Vital Statistics of the United States
1950-74 United Nations Demographic Yearbooks, 1957-74
1973-75 Central Statistical Offices of Canada, Denmark, Finland, Netherlands, Norway, Sweden and United Kingdom

PERINATAL MORTALITY RATIOS, UNITED STATES AND SELECTED COUNTRIES, 1950-75



Sources: 1950-74 U.S.: *Vital Statistics of the United States*
1950-74 *United Nations Demographic Yearbooks*
1973-75 Central Statistical Offices of Canada, Denmark, Finland, Netherlands, Norway, Sweden and United Kingdom