TOWARD A MORE COMPLETE ANALYSIS OF THE TOTAL MEAN SQUARE ERROR OF CENSUS AND SAMPLE SURVEY STATISTICS

Leroy Bailey, Bureau of the Census

Introduction

The reporting of errors associated with survey data has been generally accepted as a desirable and useful practice. However, in many cases these errors reflect only the contributions of sampling variances to the total errors of survey statistics; and the attendant assessments of the overall quality of these statistics are often based on the magnitude of this source of error. This approach to error presentation and data evaluation may sometimes lead the undiscerning data user to conclude that the only important source of survey error is that arising from sampling variability.

Survey errors may be ascribed to sampling variability, interviewers, coders, respondents, inadequate sampling frames, noninterviews, low coverage, varying interpretations of questionnaires, and other factors. At almost every juncture of the sample designing, data gathering, and processing procedures, the possibility of introducing errors exists. Candid studies of the components of the total error of survey data have and will continue to enable survey designers to (1) achieve an efficient allocation of resources among various facets of the survey; (2) improve estimates derived from survey data; and (3) more accurately assess the quality of survey data and its utility.

This paper focuses on the various components of the total mean square error of a sample mean. With the aid of an overview and some of the results of the Census Bureau efforts to assess the precision and accuracy of statistics derived from the 1970 census and one of its major sample surveys, a more detailed discussion of five of those components -- sampling variance, simple response variance, correlated response variance, nonresponse bias, and response bias, will also be provided.

Types of Survey Errors

The concept of a "true value" ($\bar{Y}_T$) of a population characteristic is included among the set of ideal goals of a statistical survey, i.e., the set of goals which would conceivably be achieved in the absence of political, social, budgetary, or procedural restrictions. A true value should be free from multiple interpretations which may lead to errors of varying sorts. More often than not, ideal goals are unobtainable, and must be replaced by a more feasible set of goals which may include an alternative population estimate ($\bar{Y}_S$), and reflect the limitations of the general survey conditions. Yet, as a result of survey errors, the actual execution of the survey produces a statistic ($\bar{y}$) which frequently differs from $\bar{Y}_S$.

Survey errors have been subdivided in various manners. Figure A illustrates a classification scheme which initially results in four major error categories: sampling variances, sampling biases, nonsampling variances, and nonsampling biases.

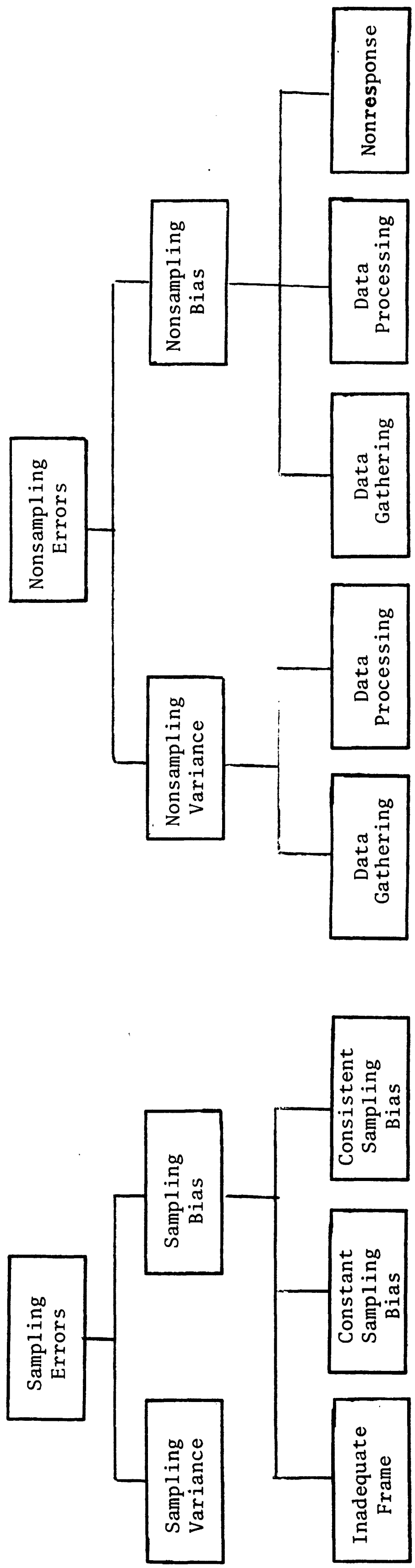

FIGURE A.--CLASSIFICATION SCHEME FOR SOURCES OF SURVEY ERRORS

Of the types of error given above, sampling variance appear to have received the greatest amount of attention. A discussion of the other forms of survey error follows.

Sampling Biases

Sampling biases can arise from many sources. Included among these are inadequate frames, and the use of certain (sometimes inappropriate) estimators of population characteristics. Some examples of these biases are the results of using the familiar ratio estimator, the simple mean to estimate an average when the sampling scheme is not simple random sampling, or a frame containing numerous duplicates and ineligible units. Some sampling biases may be easily detected and eliminated; unfortunately, others may not be so apparent, and may persist totally undetected.

Nonsampling Biases

In attempts to control such aspects of the survey design as cost, travel time, and processing difficulties, survey designers may increase the possibility of errors from other sources. If a uniform interviewing procedure is desirable, a standard questionnaire might be constructed. However, its language may present a problem to respondents far less literate than the questionnaire constructor. Perhaps the survey designer desires to achieve economies by interviewing by telephone rather than by personal interview, or by accepting responses from proxy respondents. But these innovations may stimulate responses that differ from those induced by the original survey conditions.

The potential sources of error described above belong to the class of nonsampling biases which also includes the bias of nonresponse, memory-related biases, auspices biases, and biases related to faulty record keeping. Despite the varied sources from which they may originate, many nonsampling biases can be controlled or corrected through careful survey planning and continuous evaluation.

Nonsampling Variability

Constant error tendencies which differ among a group of individuals, all of whom are assigned to perform the same survey function, such as interviewing, coding, and training, also contribute to the total variance of survey statistics. The results of census studies [14], [15], [18], and [19], suggested that the contributions of nonsampling variability are frequently large for such characteristics as unemployment and school enrollment. Despite evidence that sizeable nonsampling variances may be common to these survey statistics, they are often excluded from variance computations.

Mathematical Models

Consider a mathematical formulation developed by Tepping [12] which identifies the error components of survey statistics. There is a finite population consisting of N elements, from which a simple random sample of size n is selection. The i-th element of this population has a distribution of values y_{ij}, where the subscript j identifies the trial on which the value y_{ij} is obtained. The "true value" of a given characteristic for the i-th element is denoted by $y_{i(T)}$, and y_i represents the conditional expected value for this element over repeated trials. Additionally, let $\bar{Y}$ be the expected value of y_{ij} over all trials and the elements of all possible samples; $\bar{Y}_T$ will be the expected value of $y_{i(T)}$.

An estimate of a population parameter $\bar{Y}_S$ (which may differ from $\bar{Y}_T$) is

$$\bar{y}_j = \frac{1}{n} \sum_{i=1}^{n} y_{ij}.$$

For convenience, the subscript j will hereafter be omitted. An expression for the mean square error of the estimator is

$$\text{MSE}(\bar{y}) = E[(\bar{y} - \bar{Y}) + (\bar{Y} - \bar{Y}_S)]^2 \qquad (1)$$

$$= E(\bar{y} - \bar{Y})^2 + (\bar{Y} - \bar{Y}_S)^2 \quad . \qquad (2)$$

The first term of (2) is the total variance of the estimator, including nonsampling as well as sampling variability; the second term is the square of the net bias, which is based on the defined population parameter $\bar{Y}_S$.

Tepping decomposed this mean square error further by considering the following

$$\text{MSE}(\bar{y}) = E\{[\bar{y} - \frac{1}{n}\Sigma y_i] + [\frac{1}{n}\Sigma y_i - \frac{1}{n}\Sigma y_{i(T)}] + [\frac{1}{n}\Sigma y_{i(T)} - \bar{Y}_T] + [\bar{Y}_T - \bar{Y}_S]\}^2 \qquad (3)$$

$$= E[\bar{y} - \frac{1}{n}\Sigma y_i]^2 + E[\frac{1}{n}\Sigma y_i - \frac{1}{n}\Sigma y_{i(T)}]^2 + E[\frac{1}{n}\Sigma y_{i(T)} - \bar{Y}_T]^2 + [\bar{Y}_T - \bar{Y}_S]^2$$
$$+ 2E[\frac{1}{n}\Sigma y_i - \frac{1}{n}\Sigma y_{i(T)}][\frac{1}{n}\Sigma y_{i(T)} - \bar{Y}_T] + 2(\bar{Y}_T - \bar{Y}_S)E[\frac{1}{n}\Sigma y_i - \frac{1}{n}\Sigma y_{i(T)}]$$
$$+ 2E[\bar{y} - \frac{1}{n}\Sigma y_i][\frac{1}{n}\Sigma y_i - \frac{1}{n}\Sigma y_{i(T)}] + 2E[\bar{y} - \frac{1}{n}\Sigma y_i][\frac{1}{n}\Sigma y_{i(T)} - \bar{Y}_T] \quad . \qquad (4)$$

Under the assumption of independence among trials, equation (4) becomes:

$$\text{MSE}(\bar{y}) = \frac{\sigma_R^2}{n}[1 + \rho_R(n-1)] + \frac{N-n}{N-1}\frac{\sigma_B^2}{n} + \frac{N-n}{N-1}\frac{\sigma_S^2}{n} + \frac{N-n}{N-1}\frac{2}{n}\sigma_{BS} + \frac{2(n-1)}{n}\sigma_{RB} + \frac{2(n-1)}{n}\sigma_{RS}$$
$$+ (\bar{Y} - \bar{Y}_S)^2 - \frac{N-n}{N-1}\frac{1}{n}(\bar{Y} - \bar{Y}_T)^2. \qquad (5)$$

The terms of the above equation depict the contributions of

(1) The response variance which is comprised of a term based on the simple response variance (σ_R^2) or trial-to-trial variability in individual responses, and a term based on the correlated response variance ($\rho_R \sigma_R^2$) which reflects differences in the performances of individual coders, editors, interviewers, and other essential survey personnel;
(2) $E(y_i - y_{i(T)})^2$ or the "variance of response bias" (σ_B^2);
(3) the sampling variance (σ_S^2);
(4) the covariance of response bias and sampling deviations (σ_{BS});
(5) the covariance of response deviations and response bias (σ_{RB});
(6) the covariance of response deviations and sampling deviations.
(7) the square of the net bias $((\bar{Y} - \bar{Y}_S)^2)$; and
(8) the square of the average response bias $((\bar{Y} - \bar{Y}_T)^2)$.

Now suppose that in order to obtain estimates of a population characteristic, k interviewers are required; and each is assigned $\bar{n}$ of a total $\bar{N}$ sampling units in an interviewer assignment area. Let the survey population consist of the aggregate of the k interviewer assignment areas and $n = k\bar{n}$. The mean square error of $\bar{y}$ for a given assignment area is

$$E(\bar{y}-\bar{Y}_S)^2 = \frac{\sigma_R^2}{k\bar{n}}[1+\rho_R(\bar{n}-1)] + \frac{N-n}{N-1}\frac{\sigma_B^2}{k\bar{n}} + \frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{\sigma_S^2}{k\bar{n}} + \frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{2\sigma_{BS}}{k\bar{n}} + \frac{2(\bar{n}-1)}{k\bar{n}}\sigma_{RB} + \frac{2(\bar{n}-1)}{k\bar{n}}\sigma_{RS} + (\bar{Y}-\bar{Y}_S)^2 - \frac{\bar{N}-\bar{n}}{\bar{N}-1}\frac{1}{k\bar{n}}(\bar{Y}-\bar{Y}_T)^2. \quad (6)$$

If the expected value of the characteristic over repeated trials is substituted for the "true value",

$$\sigma_{BS} = \sigma_{RB} = \sigma_B^2 = (\bar{Y}_T - \bar{Y}_S)^2 = 0.$$

If the finite population correction is ignored, these conditions define circumstances which reduce (6) to the more familiar Hansen, Hurwitz, and Bershad model:

$$MSE(\bar{y}) = \frac{\sigma_S^2}{\bar{n}k} + \frac{\sigma_R^2}{\bar{n}k}[1 + (\bar{n}-1)\rho_R] + \frac{2(\bar{n}-1)\sigma_{RS}}{\bar{n}k} + B^2. \quad (7)$$

The Census Bureau adopted this model for utilization in large scale studies designed to estimate major components of bias and variability in census statistics. A discussion of these components follows.

Simple Response Variance

Estimates of simple or uncorrelated response variance is usually included among sampling variance estimates. Separate estimates of this source of response error are commonly derived from a replication method. This procedure is designed to reinterview all or a sample of the survey respondents through "independent" replications of survey procedures, subject to the same essential survey conditions. The estimator used with this procedure takes the following form:

$$\hat{\sigma}_R^2 = \sum_i^n \sum_j^m \frac{(x_{ij} - \bar{x}_i)^2}{n(m-1)} \quad (8)$$

where x_{ij} is the i-th response element of the j-th trial, n is the designated sample size, m is the desired number of replications, and

$$\bar{x} = \frac{1}{m}\sum_{j=1}^{m} x_{ij} .$$

The major deficiency of this estimation technique is the dissimilarity between the basic assumptions and the actual survey conditions. The assumption of independence between trials is probably never met; and frequently, modifications in personnel, procedures, and social and political attitudes invalidate the assumption of identical survey conditions.

Two CPS-Census match studies have provided estimates of simple response variances for statistics derived from the two most recent decennial censuses. The CPS or Current Population Survey is a recurring survey comprised of about 47,000 households, and is conducted by the Bureau in order to obtain essential information on employment data as well as other useful data. The 1960 match study entailed a comparison of census entries from households in the census 25 percent sample which were interviewed in the 1960 March and April CPS, to the corresponding CPS entries. In a similar manner entries from a sample of households from the 1970 census 20 percent sample were compared with the March 1970 CPS data.

Letting n represent the base for the characteristic under consideration and m=2, it can be shown that for proportions σ_R^2 as shown in equation (5) can be written as

$$\frac{1}{2n}\sum_{i=1}^{n}(x_{i1} + x_{i2} - 2x_{i1}x_{i2}) \quad (9)$$

$$= \frac{P_1 + P_2 - 2P_{12}}{2} , \quad (10)$$

where P_1 is the proportion reported in the CPS for a given characteristic, P_2 is the proportion reported in the census for the characteristic, and P_{12} is the proportion reported for both surveys. Estimates of simple response relvariance were obtained by dividing (10) by P_1^2.

Table 1 1/ provides estimates of simple response and sampling relvariances for four of the census items which were included in the 1970 match study: age, marital status, educational attainment, and employment status. The estimates of sampling relvariance are applicable to one Bernoulli trial, and under the assumption of simple random sampling are computed from the following formula

$$V_S^2 = (1=P_1)/P_1.$$

In addition, the ratio in the last column of the table gives a measure of the relative contribution of simple response variance to the sum of simple response variance and sampling variance for the given items.

The ratio of response to the sum of the simple response and sampling variances for age and marital status items was quite small. However, the corresponding ratios were rather subtantial for several of the educational attainment items and for the unemployed item.

Although the results of the CPS-Census match studies corroborated the presence of serious response variances for certain population characteristics, they should be viewed in light of the following inadequacies:

(1) Census enumerators and CPS interviewers differ. CPS interviewers participate repeatedly in training activities as well as the execution of the CPS and other sample surveys; such experiences are quite limited for census enumerators.

(2) The CPS is restricted to the noninstitutional population, and therefore is not necessarily representative of the entire country.

(3) The interviewing period for the CPS differs from that of the census; consequently the reference periods for some items, such as those listed under employment status, are not consistent. Thus, a large ratio in the last column of Table 1 may in fact reflect real change.

Correlated Response Variance

Correlated response variance describes the source of variance attributed to the correlation between response deviations of elements in the sample. If the sample size is sufficiently large, and the number of crew leaders, interviewers, coders, and editors is sufficiently small, this variance could very well dominate the total error of survey estimates.

A technique frequently used to estimate correlated response variance is the method of interpenetrating samples. It involves the splitting up of the survey sample into random groups or clusters which are approximately equal in size. This procedure permits the computation of estimates of the total variance of survey statistics: the sampling variance, and the correlated response variance. It also surmounts some of the inherent difficulties of other procedures, such as the conceivable execessive cost and difficulties in successfully duplicating experiments which are common to some replication methods.

The Census Bureau has spent a considerable amount of time investigating the correlated component of response variance, especially that part of this variance component due to interviewers. Studies which provided estimates of the variability of certain census statistics due to interviewers followed the 1960 and 1970 censuses. The two such studies were termed the 1960 Response Variance Study (RVS) which dealt with the effects of crew leaders as well as interviewers, and the 1970 Enumerator Variance Study (EVS). The 1960 experiment covered all areas of the country with the exception of certain sparsely populated or rural areas, whereas the 1970 EVS covered the decentralized mail areas, i.e., those areas in which the census was conducted by mail with enumerator follow-ups, exclusive of most highly urban areas.

Table 2 is based on a 20 percent census questionnaire that represented the decentralized mail areas. Estimates of sampling relvariances and the correlated component of response relvariances were derived from the 1970 EVS for an area of 7,500 persons. Estimates of the total survey variability within a specific cluster of interviewer assignment areas were obtained by averaging the squared deviations of interviewer means about the cluster mean. These estimates consisted of the sum of sampling variance, simple response variance and correlated response variance. The mean of squared differences within an interviewer assignment area yielded estimates of sampling variance (the sum of simple response and sampling variability) for the given cluster. Estimates of correlated response variance were then obtained by subtracting the sampling variance estimates from estimates of the total variance. The ratio in the last column of Table 2 makes a comparison of the magnitude of correlated response variance to that of sampling variance. Observe that ratios for the age items were substantially larger than those exhibited in Table 1. The marital status ratios were still fairly small, but the ratios for educational attainment and employment status items were considerable. In fact, some even exceeded 1.00.

For a more thorough discussion of the Bureau's efforts to measure the correlated response variance of census statistics, the reader is referred to references [1], [13], and [15]. A subsequent detailed report on the 1970 EVS will also be issued in the PHC(E) series.

Nonresponse Bias

Irrespective of the care with which a survey is designed, it is inevitably plagued to some extent by the problems of nonresponse bias. Nonresponse can be decomposed into two major components, noninterview and noncoverage. Noninterview refers to failure to obtain responses from elements included in the sample. It may apply to all or to only specific items of the survey. In contrast, noncoverage denotes failure to include in the sampling frame, elements of the defined survey population. Whether these sources of bias seriously affect the survey estimates depends on the extent to which the respondents and nonrespondents differ with regard to survey attributes, and the accuracy desired in survey results. For those surveys for which crude estimates are acceptable, the biases resulting from a moderate response rate may be of little importance. On the other hand, even a very high response rate can have an injurious effect on surveys requiring a high degree of accuracy.

Methods of dealing with nonresponse are divided between attempts to appreciably reduce it and efforts to adjust for its accompanying bias. The literature contains a variety of adjustment techniques and survey procedures which focus attention on survey nonresponse; they range from simple double sampling methods to complicated methods of demographic analysis. Attempts to measure the effects of nonresponse on census data have concentrated mainly on an evaluation of coverage rates for the total population and by age, race, and sex, rather than on the specific impact of nonresponse on estimates of other demographic, economic and social characteristics. However, during the period from February through June 1963 and again in September of 1965, the Census Bureau gathered data from which estimates of the CPS noninterview bias 2/ and an evaluation of the survey's noninterview adjustment procedures were made.

The CPS households which were eligible for interview and reported as noninterview cases were divided into four groups: (1) temporarily absent; (2) not-at-homes; (3) refusals; and (4) other-occupied. The 1963 study consisted of about 1,824 cases, which constituted an approximate 25 percent sample of the noninterviews during the 5 month study period. The 1965 study which excluded refusals, involved about 825 cases. These noninterview experiments required concerted efforts to "convert" or obtain responses to survey questions from all noninterview cases under study. The published CPS statistics for the time period covered by the noninterview studies were compared with statistics derived by, in effect, repeating the CPS data processing procedures so as to include data from the "converted" noninterview cases as opposed to imputes for such households. Differences between the two sets of statistics formed estimates of the CPS noninterview bias.

The findings of the CPS noninterview studies were far from being definitive because of the unsuccessful attempts to convert a satisfactory proportion of the noninterview cases. Only about 38 percent of the noninterview cases included in the 1963 study were successfully converted, while the "noninterview conversion rate" for the 1965 study was about 50 percent.

Table 3 summarizes the results of the 1965 noninterview study which pertain to labor force characteristics for the total noninstitutional population 14 years and older. Although the noninterview-related bias did not appear to be critical for the principal labor force items, it was seemingly of some consequence for some of the subcategories, the most noticeable being with a job, not at work. This item had an estimated bias that was about 5 percent of the estimated proportion for the group. Estimates of the noninterview bias and sampling variance are shown in columns 4 and 5 respectively. Column 6 gives values for the ratio of the squared noninterview bias to the sampling variance of the CPS statistics. The entries in this column suggest that for the with a job, not at work category, sampling variance is completely overshadowed by noninterview bias.

Samplers should not confuse the maintenance of relatively low noninterview rates similar to those encountered in the CPS, with the production of data which is practically unaffected by noninterview bias. Despite the limited success in following up study cases, the noninterview experiments did provide adequate reason to suspect that noninterviews may contribute substantially to the total error of some of the CPS statistics.

Response Bias

Estimates of response bias can usually be obtained through one of three popular procedures: (1) thorough administrative record checks, (2) comparisons of survey data with data secured by a separate and independent survey which is believed to be more reliable, and (3) content reinterview studies.

Content reinterview studies followed both the 1960 and 1970 censuses. After all discrepancies between the response to a specific query of the survey and the response on the reinterview had been reconciled, the reconciled response was accepted as the standard, and whenever it differed from the original survey response, a contribution was made to the overall estimate of bias for the applicable statistic.

The characteristics selected for analysis in the 1970 Content Reinterview Study included very few items which are commonly studied. Although the 1960 study provided data for a few additional characteristics which are frequently the subject of research efforts, it was still quite limited in scope and utility. Estimates of relative response bias for some of the items covered by the 1970 study appear in Table 4.

Total Error

As has already been emphasized, data assessments should be made relative to the joint effect of sampling and nonsampling errors. Table 4 shows that for the census characteristics, "Language spoken in the home". Estimates of sampling variability provided poor approximations to the total error of the given statistic. Two glaring examples of the risk involved in using sampling variance to evaluate data are shown in the ratio of the mean square error to sampling variances for "English only" and "German". These ratios are 9.45 and 7.71 respectively. In spite of these results, it should be remembered that the 1970 EVS, from which estimates of total variance were derived, was restricted to the decentralized mail areas; while the 1970 Content Reinterview Study, the source of the estimates of bias, covered the entire nation.

Again, errors whose origins are not in sampling may well be of little consequence for some surveys but to accept this condition as a general rule is quite risky. Serious errors can often occur in the absence of some form of reporting on the total error of survey statistics. The research and evaluation efforts of the Bureau of the Census have produced a considerable amount of information concerning the total error of its census and sample survey statistics. This information has consequently led to beneficial modifications in census and sample survey procedures, and to more efficient utilization of the Bureau's resources.

Summary and Recommendations

If the major objectives of a census or sample survey are to provide accurate, precise, and timely data upon which decisions, actions, and inferences are to be based, then ample consideration should be given to the coordination of existing literature, empirical data, previous work, and tested experimental procedures which are pertinent to the survey objectives. The following steps are critical to the attainment of a useful integration of pre-survey information.

(1) Preparatory Research. During the initial survey planning stages, the survey sponsor and the statistician confer on such matters as the statistical problem the survey attempts to address, and the accuracy and precision which should be connected with the various uses of the data.

After agreement has been reached on these matters, attention should be focused on the research of available literature and a review of previous experimental efforts in order to gain additional insight into the concepts, data gathering methods, and evaluation techniques relevant to the subject matter of the survey. This approach may prevent serious errors of interpretation or the repeat of experiments from which definitive inferences have already been derived.

(2) Evaluation of Valid Survey Procedures. Clearly defined survey objectives and a review of previous relevant work lay the foundation for the establishment of procedures by which the survey should be conducted. These include the hiring and training of personnel, the construction of a questionnaire, and the collection and processing of the data. Pretest and pilot studies are of inestimable worth in testing alternative questionnaires and interviewing and training techniques, and in predicting sources of potential errors. They also provide crude preliminary estimates of some of the expected sample statistics.

(3) Reporting. The interpretation and reporting of survey results should be governed by a desire to meet the objectives of the survey, and to avoid leading unsophisticated readers and users of the results to incorrect inferences. Ideally survey reporting should include the general survey findings, information about the sample design upon which the survey is based, and the presentation of survey errors. The presentation of errors should involve the representation of sampling as well as nonsampling errors and an explanation of the computation of such errors. Even if only general tables of sampling errors are included in the report, some cautionary statements regarding the existence and possible detrimental effects of nonsampling errors should also be included.

Footnotes

1/ With the exception of Table 3, the data entries for the tables found in this paper have been excerpted or compiled from data tables provided by Bailar [1].

2/ The CPS experiences on an average a monthly noninterview rate of about 4.5 percent. For a detailed discussion of the noninterview adjustment procedure employed in this survey see Part VII of Technical Paper No. 7, "The Current Population Survey--A Report on Methodology", Bureau of the Census, Washington, D.C. , 1963.

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] Cochran, W. G., Sampling Techniques, 2nd Edition, New York: John Wiley and Sons, Inc., 1963.

[3] Deming, W. E., Some Theory of Sampling, New York: Dover Publications, Inc., 1950.

[4] Fellegi, I.P., "Response Variance and Its Estimation", JASA, 59, 1964, pp. 1016-1041.

[5] Hansen, M.H., Hurwitz, W. N., and Bershad, M. A., "Measurement Errors in Censuses and Surveys", Bulletin of the International Statistical Institute , Vol. 38, No. 2, 1961, pp 359-374.

[6] Hansen, M., Hurwitz, W.N., and Madow, W.G., Sample Survey Methods and Theory Volume II, New York: John Wiley and Sons, Inc., 1953.

[7] Hansen, Morris H., Hurwitz, William N., and Pritzker, Leon, "Standardization of Procedures for the Evaluation of Data: Measurement Errors and Statistical Standards in the Bureau of the Census", Bulletin of the International Statistical Institute, 1967, 49-664.

[8] Kish, L., Survey Sampling, New York: John Wiley and Sons, Inc., 1965.

[9] Koop, J.C., "Notes for a Unified Theory of Estimation for Sample Surveys Taking into Account Response Errors", Metrika, 18, 1974, pp. 19-39.

[10] Madow, W. G., "On Some Aspects of Response Error Measurement", Proceedings of the Social Statistics Section, ASA, 1965, pp. 182-192.

[11] Palmer, S., "A Look At Alternate Imputation Procedures for CPS Noninterviews", Proceedings of the Social Statistics Section, ASA, 1967, pp. 73-80.

[12] Tepping, Benjamin J., "The Response Error Model in the Presence of "True Values", Unpublished Note, Research Center for Measurement Methods, U.S. Bureau of the Census, Washington, D.C., 1970.

[13] United States Bureau of the Census, The Post-Enumeration Survey: 1950, An Evaluation Study of the 1950 Censuses of Population and Housing, Technical Paper No. 4, Washington, D.C., 1960.

[14] United States Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing 1960: Accuracy of Data on Population Characteristics as Measured by Reinterviews, Series ER60, No.4, Washington, D.C., 1964.

[15] United States Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Effects of Interviewers and Crew Leaders, Series ER60, No.7, Washington, D.C., 1968.

[16] United States Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Record Check Study of Accuracy of Income Reporting, Series ER60, No. 8, Washington, D.C., 1970.

[17] United States Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing 1960: Accuracy of Data on Population Characteristics for Subgroups of the Total Population Series ER60, No.10, Washington, D.C., 1973.

[18] United States Bureau of the Census, Census of Population and Housing 1960: Evaluation and Research Program, Accuracy of Data for Selected Population Characteristics as Measured by Reinterviews, Series PHC(E)-9, Washington, D.C., 1974.

[19] United States Bureau of the Census, Census of Population and Housing: 1970, Evaluation and Research Program, Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match, Series PHC(E)-11, Washington, D.C., 1974.

TABLE 1.--ESTIMATED SIMPLE RESPONSE RELVARIANCES AND SAMPLING RELVARIANCES FOR SELECTED POPULATION ITEMS FOR A SIMPLE RANDOM SAMPLE OF ONE INDIVIDUAL, 1970 CENSUS

Population Characteristics	Number of persons[1/] (1)	Percent of total (2)	Relvariances: Response (3)	Relvariances: Sampling[2/] (4)	Ratio of simple response to sampling variance (5)
Age, all persons	21,502	100.0	--	--	--
Under 15 years	6,464	30.1	.03260	2.32226	.0140
15 to 24 years	3.252	15.1	.22437	5.62252	.0399
25 to 34 years	2,617	12.2	.45307	7.19672	.0630
35 to 44 years	2,585	12.0	.51998	7.33333	.0709
45 to 54 years	2,568	11.9	.48442	7.40336	.0654
55 to 64 years	1,966	9.1	.73010	9.98901	.0731
65 years and over	2,050	9.5	.44833	9.52632	.0471
Marital status, all persons 14 years and over	15,463	100.0	--	--	--
Married	10,195	65.9	.01147	.51745	.0222
Widowed	1,148	7.4	.89664	12.51351	.0717
Divorced or separated	632	4.1	4.52112	23.39024	.1933
Never married	3,488	22.6	.07538	3.42478	.0220
Educational attainment persons 25 years and over	11,780	100.0	--	--	--
Elementary 0 to 4	539	4.6	7.02065	20.73913	.3385
Elementary 5 to 7	1,025	8.7	4.96283	10.49425	.4729
Elementary 8	1,543	13.1	2.80228	6.63359	.4224
High school 1 to 3	1,974	16.8	2.01065	4.95238	.4060
High school 4	4,134	35.1	.50747	1.84900	.2745
College 1 to 3	1,241	10.5	2.71416	8.52381	.3184
College 4	802	6.8	3.51565	13.70588	.2565
College 5 or more	522	4.4	4.86712	21.72727	.2240
Employment status, persons 14 years and over	15.401	100.0	--	--	--
In labor force	8,920	57.9	.11699	.72712	.1609
Employed	8,530	55.4	.12525	.80505	.1556
Agriculture	384	2.5	10.90838	39.00000	.2797
Non-agriculture	8,146	52.9	.13040	.89036	.1465
Unemployed	390	2.5	23.88800	39.00000	.6125
Not in labor force	6,481	42.1	.22128	1.37530	.1609

[1/] Distribution is based on the CPS sample subsequent to a match with the 20-percent census sample.

[2/] Estimates were derived under the assumption of simple random sampling from a Bernoulli Distribution. This estimate includes the sampling variance.

TABLE 2.--ESTIMATED CORRELATED RESPONSE RELVARIANCES AND SAMPLING RELVARIANCES FOR SELECTED POPULATION CHARACTERISTICS FOR AN ENUMERATION BY ONE ENUMERATOR IN AN AREA OF 7,500 PERSONS, 1970 CENSUS

Population Characteristics	Number of persons 1/ (1)	Percent of total (2)	Relvariances: Response (3)	Relvariances: Sampling (4)	Ratio of correlated response to sampling variance (5)
Age, all persons	7,500	100.0	--	--	--
Under 15 years	2,228	29.7	.00038	.00203	.1872
15 to 24 years	1,185	15.8	.00000	.00455	.0000
25 to 34 years	952	12.7	.00077	.00588	.1310
35 to 44 years	900	12.0	.00010	.00543	.0184
45 to 54 years	907	12.1	.00285	.00672	.4241
55 to 64 years	668	8.9	.00553	.01013	.5459
65 years and over	308	8.8	.00397	.01033	.3843
Marital status, all persons 14 years and over	5,430	100.0	--	--	--
Married	3,551	65.4	.00000	.00060	.0000
Widowed	386	7.1	.00000	.01302	.0000
Divorced or separated	255	4.7	.00648	.02117	.3061
Never married	1,238	22.8	.00069	.00360	.1917
Educational attainment, persons 25 years and over	4.087	100.0	--	--	--
Never attended, nursery school or kindergarten	53	1.3	.24222	.10269	2.3587
Elementary 1 to 4	102	1.3	.04238	.05387	.7867
Elementary 5 to 7	323	7.9	.01590	.01662	.9567
Elementary 8	454	11.1	.01565	.01140	1.3728
High school 1 to 3	801	19.6	.00731	.00575	1.2713
High school 4	1,418	34.7	.00245	.00274	.8942
College 1 to 3	466	11.4	.00414	.01061	.3902
College 4	266	6.5	.00123	.01977	.0622
College 5 or more	200	4.9	.00704	.02729	.2580
Employment status, person 14 years and over	5,430	100.0	--	--	--
In labor force	3,133	57.7	.00021	.00060	.3500
At work	2,856	52.6	.00026	.00076	.3421
With job, not at work	92	1.7	.02551	.06042	.4222
Unemployed	130	2.4	.04543	.04179	1.0871
Armed Forces	54	1.0	.06091	.08935	.6817
Not in labor force	2,297	42.3	.00040	.00112	.3571

1/ Based on the 1970 Enumerator Variance Study which excluded very large central cities and rural areas.

TABLE 3.--LABOR FORCE CHARACTERISTICS DISTRIBUTIONS FOR POPULATION 14+, CURRENT POPULATION SURVEY, SEPTEMBER 1965

Labor Force Status	Total interviewed persons (1)	Percentage Distribution: September CPS 1/ (2)	Percentage Distribution: Noninterview study (3)	Estimates of noninterview related bias (4)	Approximate variances for estimates in (2) (5)	Ratio of squared bias to sampling variance (6)
Persons, 14 years and older	73,401	100.00	100.00	-	-	-
In Labor Force	41,208	100.00	100.00	-	-	-
Employed	39,550	95.97	96.00	.03	.01	.0900
Working 35+ Hours	30,871	75.04	75.00	.04	.04	.0400
Working 1-34 Hours	7,059	17.02	16.88	.14	.04	.4900
With a Job, Not at Work	1,620	3.92	4.12	.20	.01	4.0000
Unemployed	1,658	4.03	4.00	.03	.01	.0900
Not in Labor Force	32,193	100.00	100.00	-	-	-
Keeping House	19,162	59.57	59.50	.07	.04	.1225
School	7,217	22.38	22.18	.20	.04	1.0000
Other	4,879	15.15	15.40	.25	.04	1.5625
Unable to Work	935	2.89	2.92	.03	.01	.0900

1/ The CPS noninterview adjustment procedure essentially inflates data provided by interviewed households contained in a color-region-residence cell of a specific noninterview cluster to reflect an estimate for all sample households in the cluster.

TABLE 4.--ESTIMATES OF COMPONENTS OF MEAN SQUARE ERROR FOR "LANGUAGE SPOKEN IN HOME" FOR AN AREA OF 7,500 PERSONS

Characteristics	Number of Persons 1/ (1)	Percentage Distribution (2)	Variances: Sampling 2/ (3)	Variances: Correlated response (4)	Variances: Squared bias 3/ (5)	Variances: Mean square error (6)	Ratio of mean square error to sampling variance (7)
Language Spoken in Home							
Base (Total Persons)	7500	100.00	-	-	-	-	-
English Only	5820	77.6	3.31197	6.44328	21.55790	31.31315	9.4545
French	75	1.0	.18074	.00000	.08821	.26895	1.4880
German	218	2.9	.41730	.37870	2.42115	3.21715	7.7094
Polish	120	1.6	.26888	.20948	.06812	.54648	2.0324
Yiddish	188	2.5	.65781	.88844	.06350	1.60975	2.4471
Italian	278	3.7	.57279	1.30959	.30145	2.18383	3.8126
Spanish	142	1.9	.67175	.04068	.17140	.88380	1.3157

1/ Based on the 1970 Enumerator Variance Study which excluded the very large central cities and rural areas.
2/ Derived from the 1970 Enumerator Variance Study and includes simple response variance.
3/ Derived from the 1970 Content Reinterview Study and consist solely of response bias.