PROBLEMS AND PROSPECTS OF ENUMERATIVE DATA IN URBAN POLITICAL ANALYSIS

Peter Orleans, University of California, Los Angeles

Introduction

Where electoral behavior has been the focus of attention researchers have favored the case study approach (Lazarsfeld, et. al, 1944) or the use of survey methods (Campbell, et. al, 1960) and either by oversight or by design have tended to ignore the possibilities of comparative analyses of the interunit or intraunit type. The major exceptions are to be found in the early attempts of ecologically oriented political scientists to relate areal characteristics to individual behavior considered in the aggregate (Litchfield, 1941; Gosnell, 1936). The fallacy of the latter type of study, as W. S. Robinson has pointed out, lay in the use of the ecological correlation to make inferences about individual behavior on the basis of aggregate data (Robinson, 1951). More recently Schnore has aptly dubbed such efforts as psychological sociology (Schnore, 1961).

Now, it is increasingly possible for sociologists concerned not with individual aspects of the collective life, but rather with the analysis of structural constraints upon social organization, to revise earlier interpretations and to further develop already available techniques. Administrative units, school districts, police precincts, census tracts, political wards, land-use zones and the like, as sources of enumerative statistics constitute both a help and a hindrance to the researcher involved in such an effort. They are a help inasmuch as they provide the researcher with raw data which ordinarily cannot be obtained directly on an individual basis. Anonymity is preserved in the aggregate. And, administrative units circumscribe spatially based aggregates. They present difficulties, however, in that raw material available on an administrative unit basis carries with it the liabilities of data collected at other times, in other places, by other people, for other purposes. Neither time nor space remain constant. Thus, the boundaries of various kinds of administrative units, which rarely coincide to begin with, are often altered, and the constituents of administrative units may change.

Nevertheless, the analyst of urban political organization, cognizant of the availability of enumerative statistics collected on an administrative unit basis, has at his disposal a tantalizing array and an enormous quantity of raw material. And, what is more important, the collection of data on such a subunit basis allows for the analysis of both interunit, that is between city, as well as intraunit, or within city, variation.

In an attempt to use this material the researcher is confronted, by the nature of the data, with a series of procedural problems which have both pragmatic consequences and theoretical implications. It is the purpose of this paper to consider some of the problems involved in the use of enumerative statistics, and to note some of the assumptions which underly decisions made by the analyst in resolving such problems. This will be done by examining three specific problems encountered by the author in an analysis of the voting of selected subarea populations in two cities in the United States in elections occurring over a four year period (Orleans,1964).

Although the present discussion is limited to problems which occurred in a particular piece of research, the problems encountered are of a general nature. Thus, for instance, a method of estimating election returns for census tract populations will be considered as a specific example of the more general problem of fitting any two non-coterminous administrative units for the purpose of collating diverse sets of data. Geographic mobility and incremental growth of subarea populations as possible sources of error in the measurement of proportional participation will be examined as one example of the more general problem of anticipating and accounting for possible bias in measurement due to changes in the composition and/or constituents of administrative units. And, finally, the general problem of how to standardize measures employed in subarea analyses will be considered by examining the assumptions involved in standardizing social area indices developed by Shevky (Shevky and Bell, 1955; Bell, 1959). In short, in each case, a specific problem is employed to illustrate a more general problem.

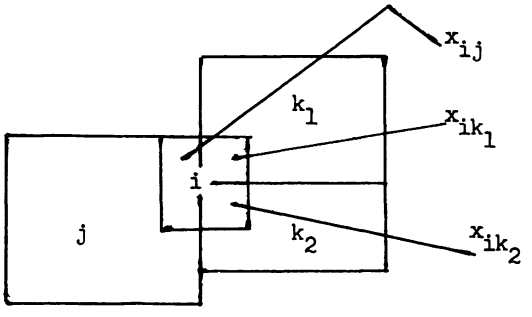
Fitting Non-Coterminous Administrative Units

As has already been suggested one of the principal uses of the various administrative units is in the collection and tabulation of enumerative statistics. It is often the case that the researcher wishes to collate data which have been organized in terms of two or more different kinds of administrative units. When the administrative units of concern are not spatially coterminous the researcher must employ some sort of fitting or estimation procedure if he is to be able to relate independent variable data available for one type of administrative unit to dependent variable data available for another. Such a procedure allows the researcher to estimate how much of the information enumerated on the basis of one type of administrative unit is to be allocated to another. The more accurate the procedure, the more reliable the results of the analysis.

Analyses of voting restricted to the use of enumerative statistics provide an appropriate example for a consideration of this problem. Data collected by the Bureau of the Census, and tabulated by census tract, provide the researcher with a wide array of independent variable (or what sociologists often refer to as facesheet) data. Attempts to relate data descriptive of census tract populations to voting data collected and tabulated on an election precinct basis

involves fitting these two types of administrative units together. Election precincts and census tracts rarely coincide. Election precincts are smaller (both in terms of geographic space and of population size) than are census tracts, and therefore it is the election precinct which must be fitted to the census tract. In other words, it is the vote summarized by election precincts which the researcher must allocate to census tracts across which election precincts fall.

When it is available, total population (though not voting age population) data enumerated on a census block basis can be used in a direct assessment of the distribution of the population in census tracts, election precincts, and areas of overlap between the two. This is because census blocks are smaller than either census tracts or election precincts, and because they do not cross the boundaries of either. Thus, an increment in accuracy is obtained by using census block data in fitting election precincts to census tracts. By doing so one does not have to make the dubious assumption that the total population of a census tract is evenly distributed in geographic space. The exact distribution of the total population can be determined on the basis of census block data. Once this distribution is determined the process of estimation involved in the fitting procedure becomes a matter of extrapolation. By making the assumption that the distribution of persons of voting age follows the known distribution of the total population, it is possible to estimate, on a non-random basis, the number of persons of voting age located in that portion of a given election precinct which overlaps the census tract of concern in the analysis. This information can be used, in the manner indicated below, to develop a weight for use in allocating election precinct data to census tract data.



i = precinct to be allocated

j = tract into which a portion of precinct i is to be allocated

k = other tracts containing a portion of precinct i (k may take values of 1, 2, 3, ... n)

x_{ij} = number of persons in precinct i and tract j

x_{ik_n} = number of persons in precinct i and tract k_n

$Y_{.j}$ = number of persons 21 years of age or older in tract j

$Y'_{.j}$ = total number of persons in tract j

$y_{.j}$ = proportion of tract j population that is 21 years of age or older; i.e., $Y_{.j}$ divided by $Y'_{.j} = y_{.j}$

$Y_{.k_n}$ = number of persons 21 years of age or older in tract k_n

$Y'_{.k_n}$ = total number of persons in tract k_n

$y_{.k_n}$ = proportion of tract k_n population that is 21 years of age or older; i.e., $Y_{.k_n}$ divided by $Y'_{.k_n}$

$t_1 = x_{ij}(y_{.j})$ the number of persons in precinct i and tract j multiplied times the proportion of the population in tract j that is 21 years of age or older; i.e., the number of persons in precinct i and tract j who are 21 years of age or older

$t_2 = \sum_{k=1}^{n} \left[(x_{ik_n})(y_{.k_n}) \right]$ the sum (overall k_n tracts) of the number of persons in precinct i and each k_n tract multiplied times the proportion of the population in each k_n tract; i.e., the number of persons in precinct i and not in tract j who are 21 years of age or older

$T = t_1 + t_2$ number of persons in precinct i who are 21 years of age or older

$M = t_1 / T$ the number of persons in precinct i and tract j who are 21 years of age or older divided by the total number of persons in precinct i (i.e., those in both tract j and the k_n tracts) who are 21 years of age or older; i.e., the proportion of persons in precinct i who are 21 years of age or older who are also in tract j or the weight used to allocate precinct i votes to tract j

x_{ij} = 511 total population in precinct i and tract j determined on the basis of block data

x_{ik_1} = 383 total population in precinct i and tract k_1 determined on the basis of block data

x_{ik_2} = 420 total population in precinct i and tract k_2 determined on the basis of block data

$Y_{.j}$ = 2257 voting age population in tract j determined on the basis of tract data

$Y'_{.j}$ = 2942 total population in tract j determined on the basis of tract data

$y_{.j}$ = .767 proportion of tract j population which is of voting age determined on the basis of computation

$Y_{.k_1}$ = 2068 voting age population in tract k_1 determined on the basis of tract data

$Y'_{.k_1}$ = 3075 total population in tract k_1 determined on the basis of tract data

$y_{.k_1}$ = .673 proportion of tract k_1 population which is of voting age determined on the basis of computation

$Y_{.k_2}$ = 743 voting age population in tract k_2 determined on the basis of tract data

$Y'_{.k_2}$ = 1124 total population in tract k_2 determined on the basis of tract data

$y_{.k_2}$ = .661 proportion of tract k_2 population which is of voting age determined on the basis of computation

t_1 = 511 (.767) = 392 estimated number of persons of voting age in precinct i and tract j

t_2 = 383 (.673) + 420 (.661) - 535 estimated number of persons of voting age in precinct i outside of tract j

T = 392 + 535 = 927 estimated number of persons of voting age in precinct i

M = 392/927 = .423 proportion of estimated voting age population in precinct i which is also estimated to be in tract j or the weight used to allocate precinct i votes to tract j; e.g., 423 of each 1000 precinct i votes are allocated to tract j

Shifts in the Composition and Constituents of Administrative Units

A second problem often encountered by researchers who use enumerative data is that of accounting for possible bias due to changes, through time, in the composition or the constituents of administrative units. Again limitations inherent in enumerative statistics employed in analyses of voting make an apt example for our consideration.

Longitudinal analyses of voting usually involve the collection of election returns at a date considerably later than the time when each election examined was held. As a result it is often the case that voter registration information is not available. This is a consequence of the widespread practice of Boards of Election of up-dating registration rolls by continually revising one list of registered voters instead of periodically developing new lists while retaining and filing old ones. Boards of Election are established to facilitate the electoral process, not research into that process and the updating procedure is most efficient for that purpose. In practice the updating method adopted by most Boards of Election means that longitudinal analyses of voting are restricted to measures of turnout indicated by ratios of persons voting to persons of voting age instead of ratios of persons voting to persons eligible to vote; eligibility simply cannot be determined.

Because the base N involved in such a measure refers to the voting age population and not to the registered population it is especially susceptible to error resulting from the inability to account for disfranchisement due to geographic mobility and the failure to meet residence requirements, as well as error due to shifts in the size of the voting age population prior to and after the date of enumeration.

The first of these sources of error, bias due to disfranchisement resulting from geographic mobility, pertains to changes in the composition (though not necessarily the constituents) of census tracts. The second source of error, shifts in the size of voting age population, is a product of a possible shift in the number of constituents. The first is a qualitative matter, the second a quantitative matter.

It is possible to use a stability coefficient to compensate for error due to differential disfranchisement. Data, available on a census tract basis, indicate (a) the total number of persons 5 years old and older as of the date of enumeration, (b) the number of persons 5 years old and older whose residence is the same at the time of enumeration as it was 5 years prior to that date, and (c) the number of persons 5 years old and older who moved to the enumerated residence from another residence in the central city of the same SMSA within the 5 year period prior to the date of enumeration. This information is sufficient to permit the construction of a gross measure of stability based on a five year interval, referring to the 5 years old and older population, and accounting for movement across SMSA boundaries. By making the assumption that movement into or out of each census tract is constant (i.e., is equivalent in each of the 5 years) a stability coefficient can be constructed, for each census tract, which would express an estimated ratio of persons with a residential tenure of one year or more to the total number of residents in the census tract. Such a ratio may be expressed as follows:

$$\frac{B + C + \{.80[A - (B + C)]\}}{A}$$

A = total persons 5 years old and older as of 1960

B = number of persons who lived in the same house in 1960 as in 1955

C = number of persons who moved to 1960 residence since 1955 and who lived in different house in the central city of the same SMSA.

Such a coefficient gives an estimate of the proportion of the population aged 5 years or over residing at the same address for one year or more. Assuming that the geographic mobility of the voting age population is roughly equivalent to that of the 5 years old and older population, the coefficient can be used to estimate the eligibility of the voting age population in terms of meeting a one year residence requirement. (On the basis of the one year criterion it was estimated that the ineligible voting age population in selected census tracts in the two cities involved in our analysis ranged from 2% to 11%).

To cope with the second limitation mentioned above, that of shifts in the size of the base population, estimates of growth and decline in the size of census tract populations would have to be made. The problem here is entirely separate from the one involved in the assessment of bias due to disfranchisement resulting from geographic mobility. It is possible to have a complete turnover in the population of a census tract without necessarily having any increment or decrement in the number of persons involved. Thus, the effect of shifts in the size of census tract populations would have to be determined by reference to previous censuses or by extrapolations into the future on the basis of past trends and current data. However, estimates over the decennial census period, for the relatively small number of persons located in each census tract, can be expected to be subject to gross distortion. Therefore, it would appear that all one can do is select elections which are temporally proximate to the date of the census enumeration, note the probable bias due to this source of error, temper the interpretation of his findings, and caution the reader about his results.

Standardization of Measures in Comparative Analyses

In any analysis involving the comparison of two or more sets of data a question is always raised as to the comparability and the generality of the measures employed. The question is inevitable inasmuch as the usual purpose of comparative research is to determine the pervasiveness of the phenomenon under consideration. To do this the researcher must be able to attribute his results, not to the procedures employed, but to the observations involved. In this sense the problem of selecting a method for standardizing measures in comparative analyses is a general question (Jaffe, 1951).

The existence of discrepancies in the assessment of positions which occur in comparative analyses of social structures first became apparent in research dealing with stratification. This occurred when researchers began asking, for example, whether the middle class in middle sized cities could be considered to be equivalent to the middle class of the metropolis. That the sociologist's concern with conceptual consequences of standardization should have occurred in the realm of stratification research is not surprising inasmuch as class has acquired the rather exalted status of a major independent variable in sociological analysis. As one observer aptly puts it, "a sociologist worth his salt, if given two basic indices of class such as income and occupation, can make a long list of predictions about the individual in question even if no further information is given." (Berger, 1963).

The problem of standardization in stratification research came to the fore because of the inability of sociologists to obtain consistent results, across communities, using reputational techniques to assess positional status. Thus, the problem came to be defined as one having to do with the instability of subjective assessments. As a result, a common resolution of the problem has been increased reliance upon objective, rather than subjective, measures of stratification. However, we would contend that even when one works exclusively with objective measures, such as occupation, education, and income, instead of subjective assessments of prestige (honor), the possibility, and therefore the problem, of discrepancies between positions in different settings (structures) remains.

Regardless of the criterion which is used as a basis for differentiating various positions, the location of positions vis a vis one another within a particular setting may depart from the relative location of positions across settings. But, the question with which one is ultimately concerned in the comparative analysis of social structures is whether differences in the relative location of positions across settings makes for differences in the organization of opportunities for various types of behavior. For example, where interest centers upon the relationship between religious affiliation and political behavior, one might be concerned with whether Catholics in a predominantly non-Catholic area are as likely to vote, or to vote Democratic, as Catholics in a predominantly Catholic area. If this were of concern, the distribution of Catholics would have to be accounted for in standardizing measures of the incidence of Catholics in various areas. Similar considerations would obtain whether the independent variable is religious affiliation, level of education, children per household, per capita income, or whatever. In other words, regardless of the form of differentiation that is the basis of classification with which one is concerned, the location of positions vis a vis one another within a given setting would seem to be a critical matter--one to be considered if a structural sociology is to be established within a comparative framework.

To illustrate the problem of standardization I will refer to a specific paradigm; social area analysis. (Shevky and Bell, 1955) I prefer to consider the problem in these terms because this particular paradigm, constructed for the purpose, already has a history of application in comparative urban research involving cities in Europe, Africa, and Latin America, as well as in the United States.

Social area analysis involves the use of indices, for example an index of social rank, which consists of the unweighted average of a number of component ratio scores; education and occupation in the case of the social rank index. In their monograph outlining computational procedures Shevky and Bell suggest that the component ratio scores should be standardized to ranges of scores empirically extant in Los Angeles in 1940 at the time of the first application of this technique of analysis. Thus, for example, the education ratio, a measure of the number of persons per 1000 (in each census tract) having elementary schooling or less, ranged from a value of 130 to one of 900, and this yields a conversion factor of .129. The standardized education ratio score for each census tract population is obtained by multiplying the difference between the score for that census tract and the lowest score, in this case 130, by the conversion factor. General acceptance of the conversion factors developed by Shevky and Bell would permit comparison of two or more census tracts, regardless of where they are located or when they are observed. The index scores can be directly related to one another because the component ratio scores involved can be referred to a common standard--the ranges of component ratio scores extant in Los Angeles in 1940.

Dissatisfaction with the standardization procedure adopted by Shevky and Bell stems from the fact that it arbitrarily anchors all subsequent analyses to the Los Angeles SMSA as defined by the Bureau of the Census in 1940, and from the conviction that there is no reason to assume that it is either theoretically representative or empirically inclusive of other urban areas.

There are several alternatives. The most common departure, in recent research, from the procedure set forth by Shevky and Bell has been the standardization of component ratio scores for a set of administrative units in a given city to their own range as of the date when the data were collected. This procedure frees each analysis from the arbitrary base established by Shevky and Bell and it preserves the autonomy of each urban area considered, but it impedes interunit, between city, comparison which cannot be accomplished without statistical interpolation. Accordingly, this standardization procedure has been adopted most often in research where concern has been with intraunit analysis, the comparison of administrative units within a given city.

A logical extension of this alternative, where interunit or between city comparisons are of concern, would be the standardization of component ratio scores to a range determined by the highest and lowest scores found among all of the administrative units in all of the cities involved in a particular analysis. In this way comparability of the numerical scores across cities would be achieved within a given analysis, but this would be done at the expense of preserving the autonomy of each urban area and therefore the importance of the interrelation of the scores within a given city. Moreover, such a standardization procedure still would not assure the comparability of component ratio scores obtained in a given analysis, with those obtained at other times, in other places, by other researchers.

A final procedure which may be considered is one in which component ratio scores are standardized to the most inclusive arbitrarily defined range of possible scores. Such a procedure would differ only in detail from the one originally proposed by Shevky and Bell. It would substitute an arbitrary, but more inclusive, range of scores for the range of scores obtained in Los Angeles in 1940. In this instance, the previously mentioned education ratio would range from 0 to 1000 and the conversion factor would be .100. A general standard will have been established permitting both interunit and intraunit comparisons, but the autonomy of specific urban areas will have been sacrificed. The numerical value of a component ratio score will always have the same meaning in absolute terms, but the relation of a given score to others in any particular configuration would remain problematic.

It has not been my intention to propose that one of the aforementioned alternatives be adopted. Rather I only want to indicate that different standardization procedures place different conceptual limitations on the data, and that such limitations and their implications must be recognized and evaluated in the process of selecting an appropriate procedure. Only future empirical research can determine how various kinds of dependent variable data are affected by the various procedures discussed above. And, eventually, when this has been done, one procedure will have to take precedence over the others for only in that way can a generic methodology for comparative urban analysis be achieved.

Conclusions

In this paper I have attempted a preliminary discussion of some of the unexpected pragmatic consequences and unrecognized theoretical implications which often result from procedural decisions in urban political research. It was suggested that the researcher should be especially cognizant of three types of problems, problems associated with the fitting of non-coterminous administrative units, problems resulting from shifts in the composition and constituents of administrative units, and problems involved in the standardization of administrative unit measures employed in comparative analyses. All of these problems require procedural decisions which must be made by the researcher who uses enumerative statistics in comparative analyses of aggregate phenomena, distributed through time and over space. As electronic data processing technology and data collection methodology become more refined the prospects for using enumerative statistics will be enhanced further, thus permitting the development of a structural sociology within a comparative framework. But, more problems of the type discussed here can be expected as well. We would suggest that such problems increasingly will make appropriate grist for the mill of statisticians interested in social science research.

References

Bell, Wendell, "Social Areas: Typology of Neighborhoods," in Marvin Sussman (ed.), Community Structure and Analysis, New York: Thomas Y. Crowell Company, 1959; pp. 61-92.

Berger, Peter, Invitation to Sociology: A Humanistic Perspective, Garden City: Doubleay Anchor Books, 1963; p. 80.

Campbell, Angus, et. al., The American Voter, New York: John Wiley and Sons, Inc., 1960.

Gosnell, Harold, Getting Out the Vote, Chicago: University of Chicago Press, 1936.

Jaffe, A. J., Handbook of Statistical Methods for Demographers, U. S. Bureau of the Census, Washington; U. S. Government Printing Office, 1951.

Litchfield, E. H., Voting Behavior in a Metropolitan Area, Ann Arbor: University of Michigan Press, 1941.

Orleans, Peter, "Partisan and Nonpartison Politics in the Metropolis: An Analysis of the Social Contexts of Political Participation," unpublished doctoral dissertation, Northwestern University, 1964.

Robinson, W. S., "Ecological Correlations and the Behavior of Individuals," American Sociological Review, 1950, pp. 315-357.

Schnore, Leo, "The Myth of Human Ecology," Sociological Inquiry, 1961, 128-139.

Shevky, Eshref, and Wendell Bell, Social Area Analysis: Theory, Illustrative Application, and Computational Procedures, Stanford: Stanford University Press: 1955.

U. S. Bureau of the Census, U. S. Censuses of Porulation and Housing: 1960. Census Tracts. Final Report PHC(1) Series.