Sampling Errors and Interpretations of Social Data Ordered in Time and Space

Frederick F. Stephan

SAMPLING ERRORS AND INTERPRETATIONS OF SOCIAL
DATA ORDERED IN TIME AND SPACE

BY FREDERICK F. STEPHAN

Census tract data are receiving increasing attention as objects of
social research, and hence the question of the applicability of formulas
of sampling to them is pertinent. But beyond this question lie others
that are of equal or greater importance.

For purely descriptive purposes these areas and the characteristics
of their populations have unquestioned and great value. When, howe-
ever, the data of census tracts are drawn upon for generalizations and
causational inferences, certain difficulties arise. The error formulas
have been applied but it is doubtful whether they are in point, any more
than they are when used upon the data of time series. Here, as in time
series, the conditions of random selection are not satisfied. One as-
sumption, that of the independence of an occurrence in one tract from
its occurrence in another (especially a contiguous one) is obviously as
false as the assumption that a price on one day is always independent
of a price on the preceding day. Data of geographic units are tied to-
gether, like bunches of grapes, not separate, like balls in an urn. Of
course mere contiguity in time and space does not of itself indicate lack
of independence between units in a relevant variable or attribute, but
in dealing with social data, we know that by virtue of their very social
character, persons, groups and their characteristics are interrelated
and not independent. Sampling error formulas may yet be developed
which are applicable to these data, but until then the older formulas
must be used with great caution. Likewise, other statistical measures
must be carefully scrutinized when applied to these data, as correla-
tion, gradients, etc.

What is most important at the present time is for the investigator to
weigh all the information obtainable, whether quantitative or not, as
it bears on the choice between alternative hypotheses. Blind applica-
tion of any methodology is especially dangerous here. Certain sugges-
tions arising out of the experience of the writer in this field may be
put forward:

(1) A vivid and realistic knowledge of the processes of urban growth
and change and of populations as distributed in both space and time
should be an exceedingly effective guide to valid generalization. This
knowledge will be itself subject to revision as the research progresses.
Differences between census tracts may reflect (a) movement of popula-

\footnote{Abstract prepared by the Chairman of this session, C. E. Gehrke.}
tion within the city, (b) influx of immigrants and migrants, (c) selective processes separating social types, such as home owner from renter, the childless from families with children, etc., (d) topographic, industrial and other features affecting desirability of areas for residential purposes, (e) epidemics, or local group phenomena, (f) differences in the average age and rate of growth of the population, (g) historical circumstances, (h) peculiarities of the census tract divisions, (i) chance fluctuations at the time of the census, and many other factors alone or in combination. The type of data under consideration will determine which of these is likely to be most significant. The selective action of economic processes is especially interesting but other selective and segregating processes should not be overlooked.

(2) Comparisons between cities should be valuable checks on generalizations drawn from census tract data. In some instances one sector of a city may be compared with other sectors. Patterns and sequences of deviations of like sign may be very significant.

(3) Census tract research will probably be most effective when considered not as a method of study complete in itself but as one step in sequence of investigations.

(4) Certain improvements in collection of data about census tracts should be made.

(5) Certain subdivisions of tract data are especially valuable in separating interwoven factors.

(6) Complexes of associated factors might be isolated and treated as wholes.

(7) Abnormal populations, such as those of business districts, institutions, non-resident groups, should be specially studied.