

RELATIONS OF SOME SOCIAL SCIENCE CONCEPTS TO STATISTICAL DATA

By: Frederick F. Stephan, Princeton University

The problem I wish to discuss with you is the alignment of statistical data and social science concepts. By this I mean the matching of a concept and a set of data related to it in such a way that one can move from data to concepts and back again in a thoroughly logical and valid manner. Unless we can do that, social theory will develop without the invigorating effect of close association with systematic observation and statistical enterprises will fail to realize their greatest potential usefulness.

What we lose if we fail to bring concepts and data into fruitful relation to each other in the social sciences may not seem very serious if we only consider past relationships. A substantial amount of social science research has utilized statistical methods and data to good effect and in turn has stimulated statistical activities. But we are only beginning. The need for statistical data in the social sciences will increase very rapidly in the future. We see a great expansion of activity in the natural sciences today, accompanied by phenomenal increases in their demands for data of great accuracy. As the social sciences mature they, too, will increase their rate of expansion and their requirements for data not only on new subjects and, with more exacting specifications, on old subjects.

Need for a New Systematic Review

Twenty-five years ago the Committee on Government Statistics and Information Services studied the problems that beset the Federal statistical agencies as they strove to meet the needs for statistical data arising out of the Great Depression and the New Deal programs. They studied these needs and made recommendations about meeting them through suitable changes in the organization of government statistical activities and the manner in which the work was performed. Ten years later, further recommendations about what should be done, who should do it, and how they should do it were made to the Hoover Commission by F. C. Mills and C. D. Long. I believe the time is approaching when there should be another systematic review of statistical activities. This time it should be directed to the relations of these activities to the research of social scientists. At the same time it should examine the extent to which the theories and concepts of social science have been developed in a manner that is conducive to the use of statistical data for their verification and further development.

This suggestion may seem to place an unnecessary burden on those of us who are struggling to sustain the standards gradually established in the past and to get work done that is needed to meet demands already recognized as most urgent. It is true that great progress has been made in organizing and supporting statistical work, in establishing standard methods and classifications, in improving comparability, in sampling, and in

processing data. It is true that there hardly seems room for additions to the program of work now being done. Nevertheless, it would be shortsighted to confine our thinking to what is practical today. If we do that, we will never find what is practical tomorrow.

Some among us will see reason to question the reasonableness of this suggestion because of the present state of the social sciences (with the possible exception of economics). It is true that there are social scientists who have no liking for statistical research. The concepts and theory that dominate social science tend to be vague, at least so they seem to a hard-headed statistician. Social scientists tend to oversimplify matters as if abstract ideas were the only reality and all the variations and exceptions we encounter when we collect data were only errors and imperfections. It is also true that they often refuse to discipline themselves to the painstaking methods we use to obtain and assemble statistics. Yet our frustrations over these shortcomings should not stop us from looking for opportunities to achieve a more effective relationship to them and their work. We have a great stake in what they do.

On our part, we should not overlook our own shortcomings. We have paid a price for the progress we have achieved in statistics. It is a certain rigidity resulting from the solution of many of our problems by somewhat arbitrary decisions and by expediency. We need not continue to limit ourselves to the solutions of problems which we adopted in the past or to confine ourselves to traditional programs, ignoring new opportunities to advance into valuable regions of work which we have neglected heretofore. I would like to make some general observations about the ways in which we have accepted restrictions on our work and relate them to some of the concepts which are important in the present and prospective development of social science.

1. The Use of Opportunistic or Pragmatic Categories

The classifications which are used in statistical work are often those which are convenient for the collection of data though they may be very inconvenient for the purposes of theoretical analysis. Many researchers accept the data gratefully since they see no chance of getting any data at all if they do not accept them in this form. For example, the concepts and theory of urban and rural differences stress the differences in values, patterns of behavior, social organization and communication, and personal development which characterize these two modes of life. As statisticians, we are baffled by these complex and elusive concepts and we take as a practical substitute a classification of political areas by the size of their populations or by population density. Obviously our ideas and theirs are not aligned in such a way that we can establish a

close or precise correspondence between them. There is an important error of translation every time we pass from one to the other. We can't continue to progress without reducing this error.

We make use of a classification of persons by socio-economic status, based on the occupation of the person or the head of his household. Social scientists attempt to make use of this classification but they can not progress until they obtain data more closely related to their concepts of stratification and class structure. They in turn must improve their concepts before they can succeed in linking them effectively to even an improved statistical version of socio-economic status.

We tend to define a child as a person under, say, eighteen years of age (or twenty-one if we are tinged with the legal definition) while social scientists look on childhood as a status established by custom which usually is defined less by age than by social attitudes. We define as a person's occupation the kind of work he spent most time doing during a week or the work at which he earned the most money. For the social scientist an occupation is a role, a part in the division of labor which is not determined by such superficial, though clear-cut, rules of classification.

We should not lightly abandon definitions which make it possible to collect data economically and accurately. The accuracy we attain, however, is spurious if it is gained at the expense of a large error of translation when the data are linked in research with social science concepts of another and incongruent character. We should free ourselves from the limitations of our previous compromises. This we can often do in special studies even when we can not do it in the regular collection of standard statistical data.

2. Dependence on Cross-sections and Current Activity

The practical exigencies of collecting data also tend to force us to take data in a brief period of time, or even of a point of time, to the neglect of changes that are going on and more gradual developments of great significance for social science theory. Even in our time series, our data often lack historical depth for the persons or groups which are important units of analysis. Thus we take the activity of a week as the basis for determining whether or not a person is in the labor force. The approach of a social scientist would be less definitive but more meaningful in taking as the basis of classifying workers and non-workers a pattern activities extending over a longer period of time. Likewise,

our data on migration tend to seize on a period of time rather than a sequence of changes. If we are to weld together the ideas of statisticians and social scientists for these kinds of human activities, we will have to find ways of revising our definitions to meet their revised concepts within the bounds of practical procedure.

3. Emphasis on Residence and the Place of Data Collection

Social scientists are greatly interested in the relations of parents and children, neighbors, relatives, school-mates and work-mates. They are concerned about community life and community institutions. Our statistics tend to ignore most interpersonal relations except some of the formal relations useful to identify individuals and a few traditional relations centered in the household. The emphasis on residence is strong. The emphasis on kinship and social interaction is weak. There are exceptions, of course, such as the collection and analysis of data for "spending units." No one can deny the great importance of the dwelling unit as a matrix of social relationships. Still we should not accept it as a sufficient substitute for the direct recording of the social relationships and activities it is intended to symbolize. Only in this way will we be able to join statistical definitions and social science concepts effectively.

These examples indicate, though they do so very inadequately, the disjuncture between the classifications which have emerged from the struggles of statisticians to improve their data and operate economically, on the one hand, and the concepts and theories developed by social scientists, on the other. Now of course some social scientists have included statistical definitions in their menu of concepts but they have not succeeded in digesting them and transforming them into the living tissues of scientific theory. We statisticians in turn have failed to assimilate into our data the more essentially social aspects of human life. Until we do, our data will fall far short of their possible usefulness for predicting and understanding, as well as for measuring, human activities and experience.

There are practical problems which warrant the expenditure of far greater sums for the collection of data than are now being spent, but only if the data fit in to the scientific processes of analysis which are needed to make them serviceable in the alleviation or solution of these problems. It is not too soon to investigate what must be done to achieve a better match of statistic and concept and to find out how it can be accomplished.