

# ANNOTATIONS

## SURVEY DESIGN AND ANALYSIS<sup>1</sup>

**T**HIS textbook is much more specialized than the title indicates. For the most part it selectively treats the analysis problems of opinion surveys. The focus is on training "analysts." These are the specialists on research teams who analyze the data but who do not necessarily either plan the survey or supervise collection of the data. The author has worked in close collaboration with the Columbia University Planning Project for Advanced Training in Social Research. One reflection of this fact is his extensive use of case materials.

The text is organized into four parts. The first part discusses "the institutional setting within which the analyst works." The large survey of today requires that the analyst use mass assembly methods and that he belong to a large research organization, which itself must seek sponsorship and subsidy. Hyman reviews some of the negative consequences of this setting such as problems of communication, excessive specialization, and conflicts over goals, methods, deadlines, and areas of responsibility.

A distinction is made between "descriptive surveys" which do not involve the study of relationships and "explanatory surveys" which do. The main problems of the descriptive survey, covered in Part II, are (1) adequate conceptualization and operational definition; (2) selection of the most "relevant" population; and (3) reduction and allowance for "response error." The distinctive problem of explanatory surveys is analyzing relationships. Procedures for this are considered in Part III.

<sup>1</sup> Hyman, Herbert: SURVEY DESIGN AND ANALYSIS. Glencoe, Ill., The Free Press, 1955, xvii + 425 pp., \$7.50.

Part IV, shorter than the others, sketches a few problems in the utilization of survey results. Most of one chapter is devoted to ways in which the opinion polls might meet some of the criticisms publicly leveled against them. In the next chapter the student is cautioned against excitement over a cross-tabulation unless the subgroups differentiated are in some way important or manipulatable. Sewell's testing of psychoanalytic theory is used to illustrate the difficulties of validating or invalidating a "discursive system" of theory.

Actually the content of Part III is the most original and merits fuller description. Analytical procedures are described which make wholesale use of subgroup comparisons. These procedures originated in work reported in *THE AMERICAN SOLDIER*. Since then several persons from the Columbia group have sought to develop a formal rationale for these procedures by distinguishing three modes of analysis, all involving the use of subgroup comparisons. Hyman and Patricia Kendall summarize this rationale in Chapter 7.

The three modes of analysis may be roughly illustrated as follows. A relationship is observed between college attendance (versus nonattendance) and later economic success. Is college attendance a genuine or spurious determinant of economic success? If spurious, then it should be possible to "wash out," or substantially reduce, this relationship by holding constant factors "extrinsic" to college attendance, such as income or occupation of father. That is, it should be possible to subgroup the total sample of college attenders and non-attenders by income or occupation of father and show that within these subgroups the original relation between college attendance and later economic success largely disappears. Typically an extrinsic factor which can explain away a relationship in this manner proves to be antecedent to the spurious determinant, just as parental occupation and income is antecedent to college attendance. On the other hand, if college attendance is a genuine determinant, then its relationship with later economic success will tend to persist within successive applications of extrinsic controls. This process of testing the genuineness of a relationship is called "explanation." There is no formal solution to the number of extrinsic factors that must be tried before accepting the

determinant as genuine. However it is felt that a thorough analysis can be distinguished from a superficial one.

Suppose that college attendance is accepted as a genuine determinant of later economic success. Then there are two broad ways to explore this relationship further. First, one may repeat the procedure above, using as controls, not factors extrinsic to college attendance, but factors integral to it or part of the same complex, such as occupational skills learned at college or contacts made there. By observing how much the relationship between college attendance and later economic success is lowered when this or that aspect of college life is held constant, one hopes to learn more about the "links" or "intervening variables" through which college attendance affects later economic success. Typically these "intervening variables" also intervene in time between determinant and effect. This mode of analysis is called "interpretation." Of course, the division between factors "extrinsic" to college attendance and integral to it, or part of the same complex, is a matter of definition and arbitrary. However several researchers working on the same problem can usually be expected to achieve a large measure of consensus.

Secondly, the economic advantage of college attendance may vary for different segments of the population. Accordingly one might classify the total sample by such factors as race, region, sex, or decade of attendance, observing how the strength of the relationship between college attendance and later economic success varies among the different subgroups. This mode of analysis is called "specification."

In all three modes of analysis the basic tool is the same—subgroup comparisons. In the first two instances the level of relationships within subgroups is compared with the original relationship taken over the entire sample. Thus the distinction between "explanation" and "interpretation" depends on the purpose of the analysis and usually too on the time sequence of determinant, "test factor," and effect. In the case of "specification" the relationships within subgroups are compared with each other, rather than with the original relation taken over the entire sample.

Throughout the book, Hyman's long experience in survey

work is obvious in the soundness of the points which he chooses to make and in his capacity to illustrate them with interesting examples. Nevertheless it must be questioned whether Hyman has produced a useful text on survey methods. The book has several serious deficiencies.

In the first place, the title *SURVEY DESIGN AND ANALYSIS* greatly overstates the ground covered. Such things as survey planning, sample design, schedule preparation, interviewer training and supervision, editing, coding, and punching are not treated. Nor are the relationships between these activities and the analysis phase systematically discussed. In particular, the consequences of sample design for analysis are ignored. Even the treatment of the analysis phase is very incomplete, being confined to a specialized set of procedures without any attempt to relate them to statistical theory.

There are pedagogical deficiencies too. The style of writing is bombastic. The ratio of abstract discussion to case materials is high in particular sections. Practice problems are offered but they are too few to represent more than a gesture.

By far the most serious deficiency, however, is a complete disregard of statistical theory. The result is that many of Hyman's discussions, if not unsound, are at least very incomplete. For example, when discussing the choice of a "relevant" population, nothing is said about having to accept a less relevant population in order to achieve an efficient sampling design, or perhaps even to attain probability sampling at all. In all his references to error and its multiple sources, no mention is made of the crucial distinction between sampling error and bias, with one of these types of error sensitive to sample size but not the other. In the section on utilizing survey results, he passes over the problem that any opinion poll, asked to collect data and make a choice between two predictions, runs a high risk of choosing the wrong alternative whenever the true division in the population is of a certain character. Thus, for example, regardless of research design, the pollster runs essentially a 50 per cent chance of being wrong about an election if the popularity of the two candidates is nearly equal.

The disregard of statistical theory is most damaging in the section concerned with the analysis of relationships. A certain

proportion of the time, sampling fluctuations will yield samples in which the relationships between a control factor and the independent and dependent variables are grossly unrepresentative of what they are in the population. This means a risk that a spurious determinant will not "wash out" when it should, as well as the converse risk that a genuine determinant will wash out even though it should not. Of course analogous risks arise in the analyses being called interpretation and specification. Unless these risks are known approximately, the results remain ambiguous and uncertified, however plausible. Hyman offers no techniques for estimating these risks. Nor is the point made that without probability sampling these risks cannot be estimated; and furthermore, that unless the sampling design is of a particularly simple sort, the estimates of these risks are much more difficult to calculate for subgroups than for the entire sample. Hyman acknowledges that the wholesale use of subgroup comparisons is often complicated by case attenuation and by "imprecision of control" (when a continuous variable is dichotomized or trichotomized in order to serve as a control). All the examples he gives have large enough samples and clear enough results so as to minimize these problems. But nothing is said about what to do in the more common case where small sample size or weak relationships render impractical an exclusive reliance on subgroup comparisons.

ROBERT G. POTTER, JR.

• • •

## IMMIGRANTS AND THEIR CHILDREN, 1850-1950<sup>1</sup>

HUTCHINSON'S IMMIGRANTS AND THEIR CHILDREN, 1850-1950, is a 1950 Census monograph. It is a namesake of Carpenter's earlier Census monograph on this subject. As described by Hutchinson:

A summary of census information on the foreign stock, up to and including the Census of 1920, is provided in the 1920 Census

<sup>1</sup> Hutchinson, E. P.: IMMIGRANTS AND THEIR CHILDREN, 1850-1950. New York, John Wiley & Sons, Inc., 1956, 391 pp. \$6.50.