CHAPTER IX

STATISTICS

The meaning of Statistics:

Statistics is conceived as a mass of figures or a collection of data. The often repeated phrase "Statistics show..." is likely to imply that a given mass of figures contains salient and unalterable characteristics that -- can easily be discerned among the mass by any person of normal intelligence. That the word statistics may apply to certain aggregates of figures is not to be denied, but that important facts contained therein are easily detected is by no means always true.

A second meaning of statistics is simply the plural of statistic, where a statistic is a - certain kind of measure used to evaluate a -- selected property of the collection of items under investigation.

A third meaning of statistics is of prime concern to us, it is the science of assembling - analyzing, characterizing, and interpreting - collections of data. In this sense, statistics is a field of study, a doctrine concerned with mathematical characterizations of aggregate of items.

Statistics, as a science, is fundamentally a branch of applied mathematics, just as - -

mechanics is mathematics applied to problems connected with bodies subjected to forces. In statistics, the applications may be made to almost any aggregate of observation or measurements. For this reason it is useful in business, economics, sociology, biology, psychology, education, physics, chemistry, agriculture, and related fields.

Definition of Population:

A population is a totality of all actual or -- conceivable objects of a certain class under - consideration. More precisely, a population - consists of numerical values connected with - these objects. Such aggregates may be finite or infinite, real or fictitious.

Definition of Sample:

A sample is a finite number of objects selec-- ted from the population. If these are chosen in such a manner that one object has as good a chance of being selected as another, we say that we have a random sample.

Random Sampling Numbers

Because of the extreme care that must someti-- mes be exercised in the selection of a sample that is truly random, statisticians frequently employ a "table of random digits". These --

digits have been produced by a mechanical or electronic process that ensures that each digit obtained is essentially independent of the digit previously obtained and is to be treated as the result of pure chance.

The Nature of Statistics

The field of statistics is extensive and varied, therefore it may be helpful to describe it in terms of certain dichotomies that characterize it.

Statistical investigations may be descriptive or inferential. Generally the former type involves fairly simple techniques; the latter demands a somewhat higher order of critical judgment and mathematical methods.

Suppose that we are confronted with a set of measurements or observations actually obtained from life. Such a set usually represents a complex of data from which it is possible to extract an almost unlimited amount of information. The task of the statistician is to select a few procedures and measures by means of which the significant aspects of the given data may be thrown into high relief. These aspects may be obtained by means of classification, graphing, and averaging. Because we are concerned only with an effective characterization of the given data themselves as they

come to us through observation, and not with estimates or conclusions involving theoretically related populations, we shall distinguish this type of statiscal analysis by means of the word descriptive. This analysis, confined exclusively to the data before us, deals with methods of recording or tabulating the constituent items, with their visual presentation, with the properties of various kinds of measures, with devices for computing them, and, in fact, with all means of giving a summary description of the data themselves.

The second type of statistical investigation is concerned with conclusions about the population that may be drawn from the data in the sample. If the data are on a large scale they may be treated as practically equivalent to the population and the properties of the sample may be considered to be like those of the population. Sometimes the sample data may constitute the population itself. If the data form a small sample drawn from the population we usually seek to derive the properties of the population from the limited information contained within the sample. It is clear that in any case we must make certain assumptions and interpret our results in the light of them. Theoretical analysis of this type is based upon the mathematical theory of probability. A second dichotomy arises in a consideration of the theoretical and the practical aspects of statistics. Mathematical or theoretical statistics seeks to derive the

laws which various populations of data and samples derived from them obey, these populations being more or less definitely specified. Applied or methodological statistics uses the theoretical results as models for the solution of practical problems. A third dichotomy occurs in considering the source of the data to be analyzed. If the data are already gathered and you have had no control over them, then you must find out what you can from what you have available. On the other hand, you may consider first the objectives of your investigation and then plan an appropriate experiment from which you can collect significant data. Here you have some control over the kind and amount of data desired. Problems connected with efficiently planned statistical experiments come under the heading of The Design of Experiments. Large sample methods are most useful when the data are already in; small sample techniques predominate when you gather information from an appropriate experiment.

The data of statistics consist essentially of numerical values derived from measurements, observations, interviews, experiments, and so on-values that, in a given set, are generally different and unpredictable. Such magnitudes are called random or chance variables, because many chance causes operate to produce them. The theory of probability deals with such variables and is the cornerstone of mathematical statistics. The word variate is often used to

denote a numerical outcome that is the result of chance.

Some Nonmathematical Aspects

Although statistics is fundamentally a mathematical study, there are certain phases of it that may be termed nonmathematical, and that are, in many respects, prerequisite to successful statistical analysis. It is useless to perform elaborate calculations and to derive conclusions from data of questioned reliability and ambiguous meaning. (1) Sources. Data may arise from books, reports, tables, and so forth, or from the results of original investigation, such as interviews, questionnaires, direct observations, actual measurements, and the like. (2) Definitions. The nature of the attributes studied in a statistical investigation may profoundly influence the magnitudes involved and the conclusions drawn; for this reason it is essential that basic terms be clearly understood. It is often necessary to make precise definitions, or to describe in detail the method of measurement. (3) Purposes. The nature and extent of a statistical study are determined largely by the purpose in view. Frequently, the objective is a single concise result displayed in a manner readily understood by "the man on the street". In such cases, finely drawn distinctions, precise mathematical results, and distracting details may be sacrificed to a simple, graphic representa--

tion that catches the eye and impresses the mind. Other investigations may aim to establish a thesis or to destroy one-by means - of a comprehensive analysis of all relevant factors. Such studies often require elaborate technical equipment on the part of the reader as well as the writer.

I N D E X