STATISTICS

Statistics is a group of methods used to collect, analyze, present, and interpret data and to make decisions. It is the art of learning from data. It is concerned with the collection of data, their subsequent description, and their analysis, which often leads to the drawing of conclusions. Every day we make decisions that may be personal, business related, or of some other kind. Usually these decisions are made under conditions of uncertainty. Many times, the situations or problems we face in the real world have no precise or definite solution. Statistical methods help us make scientific and intelligent decisions in such situations. Decisions made by using statistical methods are called *educated guesses.* Decisions made without using statistical (or scientific) methods are *pure guesses* and, hence, may prove to be unreliable. For example, opening a large store in an area with or without assessing the need for it may affect its success. Like almost all fields of study, statistics has two aspects: theoretical and applied. *Theoretical* or *mathematical statistics* deals with the development, derivation, and proof of statistical theorems, formulas, rules, and laws. *Applied statistics* involves the applications of those theorems, formulas, rules, and laws to solve real-world problems. This text is concerned with applied statistics and not with theoretical statistics.

Statistics enables us to summarize lots of information about people, for example after standardization when a lot of people have been measured on some variable and their scores collected. The test-maker needs to understand the distribution of raw scores among the standardization sample. Every person will have a score and the overall data set should be investigated in a meaningful way. The best way of doing this is to work out how often a score is obtained by different people, giving us what is called its frequency, and then to draw a graph summarizing all of the frequencies. For example, I might ask a group of students how many cups of coffee they drink each day and the data obtained could be:

1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 4 55 5 5 5 5 5 5 5 6 6 6 6 6 6 6 7 7 7
7 7 7 8 8 8 8 9 9

Investigation of the data suggests that two people drank one cup a day, four drank two cups, six drank three cups, seven drank four, and so on.

The Five Basic Words of Statistics
The five words *population*, *sample*, *parameter*, *statistic* (singular), and *variable* form the basic vocabulary of statistics. You cannot learn much about statistics unless you first learn the meanings of these five words.

Population
CONCEPT All the members of a group about which you want to draw a conclusion.
EXAMPLES All Indian citizens who are currently registered to vote, all patients treated at a particular hospital last year, the entire daily output of a cereal factory's production line.

Sample
CONCEPT The part of the population selected for analysis.

EXAMPLES The registered voters selected to participate in a recent survey concerning their intention to vote in the next election, the patients selected to fill out a patient-satisfaction questionnaire, 100 boxes of cereal selected from a factory's production line.

Parameter
CONCEPT A numerical measure that describes a characteristic of a population.

EXAMPLES The percentage of all registered voters who intend to vote in the next election, the percentage of all patients who are very satisfied with the care they received, the average weight of all the cereal boxes produced on a factory's production line on a particular day.

Statistic
CONCEPT A numerical measure that describes a characteristic of a sample.

EXAMPLES The percentage in a sample of registered voters who intend to vote in the next election, the percentage in a sample of patients who are very satisfied with the care they received, the average weight of a sample of cereal boxes produced on a factory's production line on a particular day.

INTERPRETATION Calculating statistics for a sample is the most common activity, because collecting population data is impractical for most actual decision-making situations.

Variable
CONCEPT A characteristic of an item or an individual that will be analyzed using statistics.

EXAMPLES Gender, the household income of the citizens who voted in the last presidential election, the publishing category (hardcover, trade paperback, mass-market paperback, textbook) of a book, the number of varieties
of a brand of cereal.

INTERPRETATION All the variables taken together form the data of an analysis. Although you may have heard people saying that they are analyzing their data, they are, more precisely, analyzing their variables. You should distinguish between a variable, such as gender, and its **value** for an individual, such as male. An **observation** is all the values for an individual item in the sample. For example, a survey might contain two variables, gender and age. The first observation might be male, 40. The second observation might be female, 45. The third observation might be female, 55. A **variable** is sometimes known as a column of data because of the convention of entering each observation as a unique row in a table of data.

The Branches of Statistics
Two branches, *descriptive statistics* and *inferential statistics*, comprise the field of statistics.

Descriptive Statistics
CONCEPT The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

EXAMPLES The average age of citizens who voted for the winning candidate in the last presidential election, the average length of all books about statistics, the variation in the weight of 100 boxes of cereal selected from a factory's production line.

INTERPRETATION You are most likely to be familiar with this branch of statistics, because many examples arise in everyday life. Descriptive statistics forms the basis for analysis and discussion in such diverse fields as securities trading, the social sciences, government, the health sciences, and professional sports. A general familiarity and widespread availability of descriptive methods in many calculating devices and business software can often make using this branch of statistics seem deceptively easy.

Inferential Statistics
CONCEPT The branch of statistics that analyzes sample data to draw conclusions about a population.

EXAMPLE A survey that sampled 2,001 full- or part-time workers ages 50 to 70, conducted by the American Association of Retired Persons (AARP), discovered that 70% of those polled planned to work past the traditional mid- 60s retirement age.

INTERPRETATION When you use inferential statistics, you start with a hypothesis and look to see whether the data are consistent with that hypothesis. Inferential statistical methods can be easily misapplied or misconstrued, and many inferential methods require the use of a calculator or computer.