

Objective 2

Differentiate between the two branches of statistics.

Historical Note

The origin of descriptive statistics can be traced to data collection methods used in censuses taken by the Babylonians and Egyptians between 4500 and 3000 B.C. In addition, the Roman Emperor Augustus (27 B.C.–A.D. 17) conducted surveys on births and deaths of the citizens of the empire, as well as the number of livestock each owned and the crops each citizen harvested yearly.

Historical Note

Inferential statistics originated in the 1600s, when John Graunt published his book on population growth, *Natural and Political Observations Made upon the Bills of Mortality*. About the same time, another mathematician/astronomer, Edmund Halley, published the first complete mortality tables. (Insurance companies use mortality tables to determine life insurance rates.)

Data can be used in different ways. The body of knowledge called statistics is sometimes divided into two main areas, depending on how data are used. The two areas are

1. Descriptive statistics
2. Inferential statistics

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

In *descriptive statistics* the statistician tries to describe a situation. Consider the national census conducted by the U.S. government every 10 years. Results of this census give you the average age, income, and other characteristics of the U.S. population. To obtain this information, the Census Bureau must have some means to collect relevant data. Once data are collected, the bureau must organize and summarize them. Finally, the bureau needs a means of presenting the data in some meaningful form, such as charts, graphs, or tables.

The second area of statistics is called *inferential statistics*.

Inferential statistics consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships among variables, and making predictions.

Here, the statistician tries to make inferences from *samples* to *populations*. Inferential statistics uses **probability**, i.e., the chance of an event occurring. You may be familiar with the concepts of probability through various forms of gambling. If you play cards, dice, bingo, or lotteries, you win or lose according to the laws of probability. Probability theory is also used in the insurance industry and other areas.

It is important to distinguish between a sample and a population.

A **population** consists of all subjects (human or otherwise) that are being studied.

Most of the time, due to the expense, time, size of population, medical concerns, etc., it is not possible to use the entire population for a statistical study; therefore, researchers use samples.

A **sample** is a group of subjects selected from a population.

If the subjects of a sample are properly selected, most of the time they should possess the same or similar characteristics as the subjects in the population. The techniques used to properly select a sample will be explained in Section 1–3.

An area of inferential statistics called **hypothesis testing** is a decision-making process for evaluating claims about a population, based on information obtained from samples. For example, a researcher may wish to know if a new drug will reduce the number of heart attacks in men over 70 years of age. For this study, two groups of men over 70 would be selected. One group would be given the drug, and the other would be given a placebo (a substance with no medical benefits or harm). Later, the number of heart attacks occurring in each group of men would be counted, a statistical test would be run, and a decision would be made about the effectiveness of the drug.

Statisticians also use statistics to determine *relationships* among variables. For example, relationships were the focus of the most noted study in the 20th century, “Smoking and Health,” published by the Surgeon General of the United States in 1964. He stated that after reviewing and evaluating the data, his group found a definite relationship between smoking and lung cancer. He did not say that cigarette smoking actually causes lung cancer, but that there is a relationship between smoking and lung cancer. This conclusion was based on a study done in 1958 by Hammond and Horn. In this study, 187,783 men were observed over a period of 45 months. The death rate from

Speaking of Statistics

Statistics and the New Planet

In the summer of 2005, astronomers announced the discovery of a new planet in our solar system. Astronomers have dubbed it Eris. They also discovered that it has a moon that is larger than Pluto. Eris is about 9 billion miles from the Sun. (Some sources say 10 billion.) Its diameter is about 1600 miles. Its surface temperature has been estimated at -400°F , and it takes 560 years to circle the Sun.

How does Eris compare to the other planets? Let's look at the statistics.



Planet	Diameter (miles)	Distance from the Sun (millions of miles)	Orbital period (days)	Mean temperature ($^{\circ}\text{F}$)	Number of moons
Mercury	3,032	36	88	333	0
Venus	7,521	67.2	224.7	867	0
Earth	7,926	93	365.2	59	1
Mars	4,222	141.6	687	-85	2
Jupiter	88,846	483.8	4,331	-166	63
Saturn	74,897	890.8	10,747	-220	62
Uranus	31,763	1,784.8	30,589	-320	27
Neptune	30,775	2,793.1	59,800	-330	13
Pluto	1,485	3,647.2	90,588	-375	3

Source: NASA.

With these statistics, we can make some comparisons. For example, Eris is about the size of the planet Mars, but its volume is only 25% greater than Pluto's. (Compare the volumes.) It takes about twice as long to circle the Sun as Pluto. What other comparisons can you make?

Unusual Stat

Twenty-nine percent of Americans want their boss's job.

lung cancer in this group of volunteers was 10 times as great for smokers as for nonsmokers.

Finally, by studying past and present data and conditions, statisticians try to make predictions based on this information. For example, a car dealer may look at past sales records for a specific month to decide what types of automobiles and how many of each type to order for that month next year.

Applying the Concepts 1-1

Attendance and Grades

Read the following on attendance and grades, and answer the questions.

A study conducted at Manatee Community College revealed that students who attended class 95 to 100% of the time usually received an A in the class. Students who attended class

Unusual Stat

Only one-third of crimes committed are reported to the police.

80 to 90% of the time usually received a B or C in the class. Students who attended class less than 80% of the time usually received a D or an F or eventually withdrew from the class.

Based on this information, attendance and grades are related. The more you attend class, the more likely it is you will receive a higher grade. If you improve your attendance, your grades will probably improve. Many factors affect your grade in a course. One factor that you have considerable control over is attendance. You can increase your opportunities for learning by attending class more often.

1. What are the variables under study?
2. What are the data in the study?
3. Are descriptive, inferential, or both types of statistics used?
4. What is the population under study?
5. Was a sample collected? If so, from where?
6. From the information given, comment on the relationship between the variables.

See page 33 for the answers.

1-2**Variables and Types of Data****Objective 3**

Identify types of data.

As stated in Section 1-1, statisticians gain information about a particular situation by collecting data for random variables. This section will explore in greater detail the nature of variables and types of data.

Variables can be classified as qualitative or quantitative. **Qualitative variables** are variables that can be placed into distinct categories, according to some characteristic or attribute. For example, if subjects are classified according to gender (male or female), then the variable *gender* is qualitative. Other examples of qualitative variables are religious preference and geographic locations.

Quantitative variables are numerical and can be ordered or ranked. For example, the variable *age* is numerical, and people can be ranked in order according to the value of their ages. Other examples of quantitative variables are heights, weights, and body temperatures.

Quantitative variables can be further classified into two groups: discrete and continuous. *Discrete variables* can be assigned values such as 0, 1, 2, 3 and are said to be countable. Examples of discrete variables are the number of children in a family, the number of students in a classroom, and the number of calls received by a switchboard operator each day for a month.

Discrete variables assume values that can be counted.

Continuous variables, by comparison, can assume an infinite number of values in an interval between any two specific values. Temperature, for example, is a continuous variable, since the variable can assume an infinite number of values between any two given temperatures.

Continuous variables can assume an infinite number of values between any two specific values. They are obtained by measuring. They often include fractions and decimals.

The classification of variables can be summarized as follows:

