1.1 Statistics and Types of Statistics

In this section we will learn about statistics and types of statistics.

1.1.1 What Is Statistics?

The word **statistics** has two meanings. In the more common usage, *statistics* refers to numerical facts. The numbers that represent the income of a family, the age of a student, the percentage of passes completed by the quarterback of a football team, and the starting salary of a typical college graduate are examples of statistics in this sense of the word. A 1988 article in *U.S. News & World Report* mentioned that "Statistics are an American obsession."[1] During the 1988 baseball World Series between the Los Angeles Dodgers and the Oakland A's, the then NBC commentator Joe Garagiola reported to the viewers numerical facts about the players' performances. In response, fellow commentator Vin Scully said, "I love it when you talk statistics." In these examples, the word *statistics* refers to numbers.

The following examples present some statistics:

1. During March 2014, a total of 664,000,000 hours were spent by Americans watching March Madness live on TV and/or streaming (*Fortune* Magazine, March 15, 2015).
2. Approximately 30% of Google's employees were female in July 2014 (*USA TODAY*, July 24, 2014).
3. According to an estimate, an average family of four living in the United States needs $130,357 to live the American dream (*USA TODAY*, July 7, 2014).
4. Chicago's O'Hare Airport was the busiest airport in 2014, with a total of 881,933 flight arrivals and departures.
5. In 2013, author James Patterson earned $90 million from the sale of his books (*Forbes*, September 29, 2014).
6. About 22.8% of U.S. adults do not have a religious affiliation (*Time*, May 25, 2015).
7. Yahoo CEO Marissa Mayer was the highest paid female CEO in America in 2014, with a total compensation of $42.1 million.

The second meaning of *statistics* refers to the field or discipline of study. In this sense of the word, *statistics* is defined as follows.

Statistics **Statistics** is the science of collecting, analyzing, presenting, and interpreting data, as well as of making decisions based on such analyses.

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.

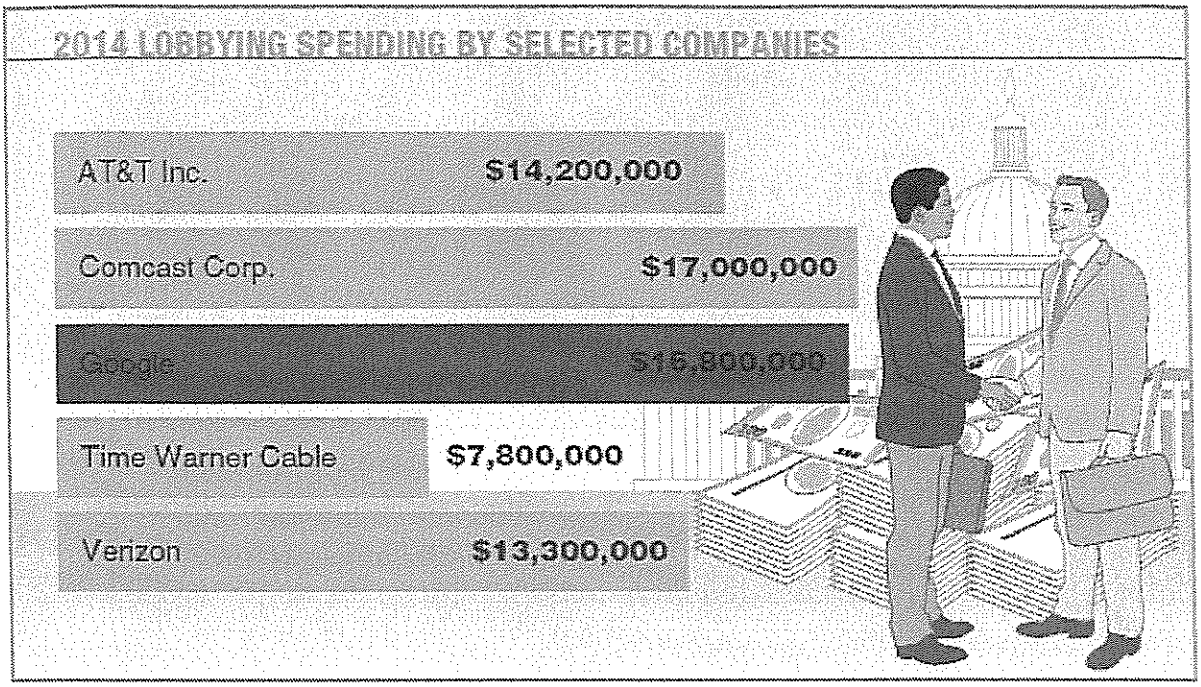
Like almost all fields of study, statistics has two aspects: theoretical and applied. *Theoretical* or *mathematical statistics* deals with the development, derivation, and proof of statistical theorems, formulas, rules, and laws. *Applied statistics* involves the applications of those theorems, formulas, rules, and laws to solve real-world problems. This text is concerned with applied statistics and not with theoretical statistics. By the time you finish studying this book, you will have learned how to think statistically and how to make educated guesses.

1.1.2 Types of Statistics

Broadly speaking, applied statistics can be divided into two areas: **descriptive statistics** and **inferential statistics**.

CASE STUDY 1-1

2014 LOBBYING SPENDING BY SELECTED COMPANIES



Data source: Fortune Magazine, June 1, 2015

The accompanying chart shows the lobbying spending by five selected companies during 2014. Many companies spend millions of dollars to win favors in Washington. According to *Fortune* Magazine of June 1, 2015, "Comcast has remained one of the biggest corporate lobbyists in the country." In 2014, Comcast spent $17 million, Google spent $16.8 million, AT&T spent $14.2 million, Verizon spent $13.3 million, and Time Warner Cable spent $7.8 million on lobbying. These numbers simply describe the total amounts spent by these companies on lobbying. We are not drawing any inferences, decisions, or predictions from these data. Hence, this data set and its presentation is an example of descriptive statistics.

Descriptive Statistics

Suppose we have information on the test scores of students enrolled in a statistics class. In statistical terminology, the whole set of numbers that represents the scores of students is called a **data set**, the name of each student is called an **element**, and the score of each student is called an **observation**. (These terms are defined in more detail in Section 1.2.)

Many data sets in their original forms are usually very large, especially those collected by federal and state agencies. Consequently, such data sets are not very helpful in

drawing conclusions or making decisions. It is easier to draw conclusions from summary tables and diagrams than from the original version of a data set. So, we summarize data by constructing tables, drawing graphs, or calculating summary measures such as averages. The portion of statistics that helps us do this type of statistical analysis is called **descriptive statistics**.

Descriptive Statistics **Descriptive statistics** consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures.

Chapters 2 and 3 discuss descriptive statistical methods. In Chapter 2, we learn how to construct tables and how to graph data. In Chapter 3, we learn how to calculate numerical summary measures, such as averages.

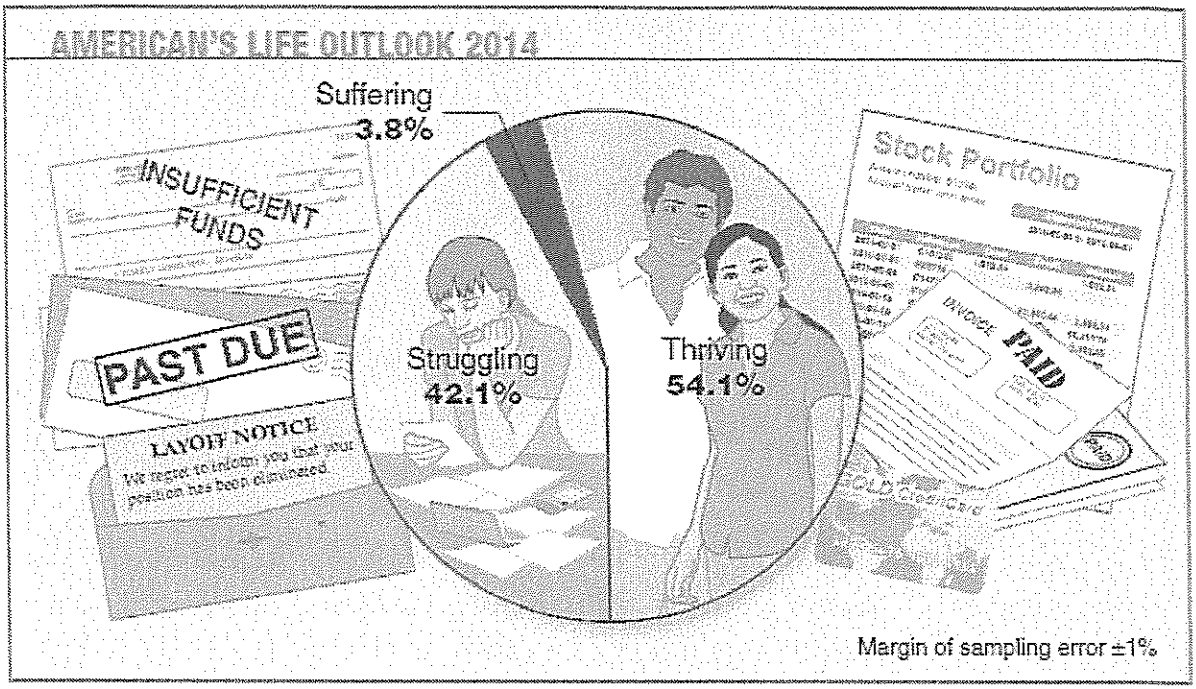
Case Study 1-1 presents an example of descriptive statistics.

Inferential Statistics

In statistics, the collection of all elements of interest is called a **population**. The selection of a portion of the elements from this population is called a **sample**. (Population and sample are discussed in more detail in Section 1.5.)

CASE STUDY 1-2

AMERICANS' LIFE OUTLOOK, 2014



Data source: Gallup-Healthways Well-Being Index

A poll of 176,903 American adults, aged 18 and older, was conducted January 2 to December 30, 2014, as part of the Gallup-Healthways Well-Being Index survey. Gallup and Healthways have been "tracking Americans' life evaluations daily" since 2008. According to this poll, in 2014, Americans' outlook on life was the best in seven years, as 54.1% "rated their lives highly enough to be considered thriving," 42.1% said they were struggling, and 3.8% mentioned that they were suffering. As mentioned in the chart, the margin of sampling error was ± 1%. In Chapter 8, we will discuss the concept of margin of error, which can be combined with these percentages when making inferences. As we notice, the results described in the chart are obtained from a poll of 176,903 adults. We will learn in later chapters how to apply these results to the entire population of adults. Such decision making about the population based on sample results is called inferential statistics.

A major portion of statistics deals with making decisions, inferences, predictions, and forecasts about populations based on results obtained from samples. For example, we may make some decisions about the political views of all college and university students based on the political views of 1000 students selected from a few colleges and universities. As another example, we may want to find the starting salary of a typical college graduate. To do so, we may select 2000 recent college graduates, find their

starting salaries, and make a decision based on this information. The area of statistics that deals with such decision-making procedures is referred to as **inferential statistics**. This branch of statistics is also called *inductive reasoning* or *inductive statistics*.

Inferential Statistics **Inferential statistics** consists of methods that use sample results to help make decisions or predictions about a population.

Case Study 1-2 presents an example of inferential statistics. It shows the results of a survey in which American adults were asked about their opinions about their lives.

Chapters 8 through 15 and parts of Chapter 7 deal with inferential statistics.

Probability, which gives a measurement of the likelihood that a certain outcome will occur, acts as a link between descriptive and inferential statistics. Probability is used to make statements about the occurrence or nonoccurrence of an event under uncertain conditions. Probability and probability distributions are discussed in Chapters 4 through 6 and parts of Chapter 7.

EXERCISES

CONCEPTS AND PROCEDURES

1.1 Briefly describe the two meanings of the word *statistics*.

1.2 Briefly explain the types of statistics.

APPLICATIONS

1.3 Which of the following is an example of descriptive statistics and which is an example of inferential statistics? Explain.

a. In a survey by *Fortune* Magazine and SurveyMonkey, participants were asked what was the most important factor when purchasing groceries (*Fortune*, June 1, 2015). The following table lists the summary of the responses of these participants. Assume that the maximum margin of error is ± 1.5%.

Factor	Percent of Respondents
Price	42.4
Nutrition	36.0
Absence of additives	16.4
Number of calories	3.8
Carbon footprint	1.5

b. The following table gives the earnings of the world's top seven female professional athletes for the year 2014 (ceoworld.biz).

Female Professional Athlete	2014 Earnings (millions of dollars)
Maria Sharapova	24.4
Li Na	23.6
Serena Williams	22.0
Kim Yuna	16.3
Danica Patrick	15.0
Victoria Azarenka	11.1
Caroline Wozniacki	10.8

1.2 Basic Terms

It is very important to understand the meaning of some basic terms that will be used frequently in this text. This section explains the meaning of an element (or member), a variable, an observation, and a data set. An element and a data set were briefly defined in Section 1.1. This section defines these terms formally and illustrates them with the help

of an example.

Table 1.1 gives information, based on *Forbes* magazine, on the total wealth of the world's eight richest persons as of March 2015. Each person listed in this table is called an **element** or a **member** of this group. Table 1.1 contains information on eight elements. Note that elements are also called **observational units**.

Table 1.1 Total Wealth of the World's Eight Richest Persons

Name	Total Wealth (billions of dollars)
Bill Gates	79.2
Carlos Slim Helu	77.1
Warren Buffett	72.7
Amancio Ortega	64.5
Larry Ellison	54.3
Charles Koch	42.9
David Koch	42.9
Christy Walton	41.7

Source: *Forbes*, March 23, 2015.

Element or Member An **element** or **member** of a sample or population is a specific subject or object (for example, a person, firm, item, state, or country) about which the information is collected.

The total wealth in our example is called a variable. The total wealth is a characteristic of these persons on which information is collected.

Variable A **variable** is a characteristic under study that assumes different values for different elements. In contrast to a variable, the value of a *constant* is fixed.

A few other examples of variables are household incomes, the number of houses built in a city per month during the past year, the makes of cars owned by people, the gross profits of companies, and the number of insurance policies sold by a salesperson per day during the past month.

In general, a variable assumes different values for different elements, as illustrated by the total wealth for the eight persons in Table 1.1. For some elements in a data set, however, the values of the variable may be the same. For example, if we collect information on incomes of households, these households are expected to have different incomes, although some of them may have the same income.

A variable is often denoted by x, y, or z. For instance, in Table 1.1, the total wealth for persons may be denoted by any one of these letters. Starting with Section 1.7, we will begin to use these letters to denote variables.

Each of the values representing the total wealths of the eight persons in Table 1.1 is called an **observation** or **measurement**.

Observation or Measurement The value of a variable for an element is called an **observation** or **measurement**.

From Table 1.1, the total wealth of Warren Buffett was \$72.7 billion. The value \$72.7 billion is an observation or a measurement. Table 1.1 contains eight observations, one for each of the eight persons.

The information given in Table 1.1 on the total wealth of the eight richest persons is called the data or a **data set**.

Data Set A **data set** is a collection of observations on one or more variables.

Other examples of data sets are a list of the prices of 25 recently sold homes, test scores of 15 students, opinions of 100 voters, and ages of all employees of a company.

EXERCISES

CONCEPTS AND PROCEDURES

1.4 Explain the meaning of an element, a variable, an observation, and a data set.

APPLICATIONS

1.5 The following table lists the number of deaths by cause as reported by the Centers for Disease Control and Prevention on February 6, 2015 (*Source*: www.cdc.gov).

Cause of Death	Number of Deaths
Heart disease	611,105
Cancer	584,881
Accidents	130,557
Stroke	128,978
Alzheimer's disease	84,767
Diabetes	75,578
Influenza and Pneumonia	56,979
Suicide	41,149

Briefly explain the meaning of a member, a variable, a measurement, and a data set with reference to the information in this table.

1.6 The following table lists the number of deaths by cause as reported by the Centers for Disease Control and Prevention on February 6, 2015 (*Source*: www.cdc.gov).

Cause of Death	Number of Deaths
Heart disease	611,105
Cancer	584,881
Accidents	130,557
Stroke	128,978
Alzheimer's disease	84,767
Diabetes	75,578
Influenza and Pneumonia	56,979
Suicide	41,149

a. What is the variable for this data set?

b. How many observations are in this data set?

c. How many elements does this data set contain?

1.3 Types of Variables

In Section 1.2, we learned that a variable is a characteristic under investigation that assumes different values for different elements. Family income, height of a person, gross sales of a company, price of a college textbook, make of the car owned by a family, number of accidents, and status (freshman, sophomore, junior, or senior) of a student enrolled at a university are examples of variables.

A variable may be classified as quantitative or qualitative. These two types of variables are explained next.

1.3.1 Quantitative Variables

Some variables (such as the price of a home) can be measured numerically, whereas others (such as hair color) cannot. The price of a home is an example of a **quantitative variable** while hair color is an example of a **qualitative variable**.

Quantitative Variable A variable that can be measured numerically is called a **quantitative variable**. The data collected on a quantitative variable are called **quantitative data**.

Income, height, gross sales, price of a home, number of cars owned, and number of accidents are examples of quantitative variables because each of them can be expressed numerically. For instance, the income of a family may be $81,520.75 per year, the gross sales for a company may be $567 million for the past year, and so forth. Such quantitative variables may be classified as either *discrete variables* or *continuous variables*.

Discrete Variables

The values that a certain quantitative variable can assume may be countable or noncountable. For example, we can count the number of cars owned by a family, but we cannot count the height of a family member, as it is measured on a continuous scale. A variable that assumes countable values is called a **discrete variable**. Note that there are no possible intermediate values between consecutive values of a discrete variable.

Discrete Variable A variable whose values are countable is called a **discrete variable**. In other words, a discrete variable can assume only certain values with no intermediate values.

For example, the number of cars sold on any given day at a car dealership is a discrete variable because the number of cars sold must be 0, 1, 2, 3, ... and we can count it. The number of cars sold cannot be between 0 and 1, or between 1 and 2. Other examples of discrete variables are the number of people visiting a bank on any day, the number of cars in a parking lot, the number of cattle owned by a farmer, and the number of students in a class.

Continuous Variables

Some variables assume values that cannot be counted, and they can assume any numerical value between two numbers. Such variables are called **continuous variables**.

Continuous Variable A variable that can assume any numerical value over a certain interval or intervals is called a **continuous variable**.

The time taken to complete an examination is an example of a continuous variable because it can assume any value, let us say, between 30 and 60 minutes. The time taken may be 42.6 minutes, 42.67 minutes, or 42.674 minutes. (Theoretically, we can measure time as precisely as we

want.) Similarly, the height of a person can be measured to the tenth of an inch or to the hundredth of an inch. Neither time nor height can be counted in a discrete fashion. Other examples of continuous variables are the weights of people, the amount of soda in a 12-ounce can (note that a can does not contain exactly 12 ounces of soda), and the yield of potatoes (in pounds) per acre. Note that any variable that involves money and can assume a large number of values is typically treated as a continuous variable.

1.3.2 Qualitative or Categorical Variables

Variables that cannot be measured numerically but can be divided into different categories are called **qualitative** or **categorical variables**.

Qualitative or Categorical Variable A variable that cannot assume a numerical value but can be classified into two or more nonnumeric categories is called a **qualitative** or **categorical variable**. The data collected on such a variable are called **qualitative data**.

For example, the status of an undergraduate college student is a qualitative variable because a student can fall into any one of four categories: freshman, sophomore, junior, or senior. Other examples of qualitative variables are the gender of a person, the make of a computer, the opinions of people, and the make of a car.

Figure 1.1 summarizes the different types of variables.

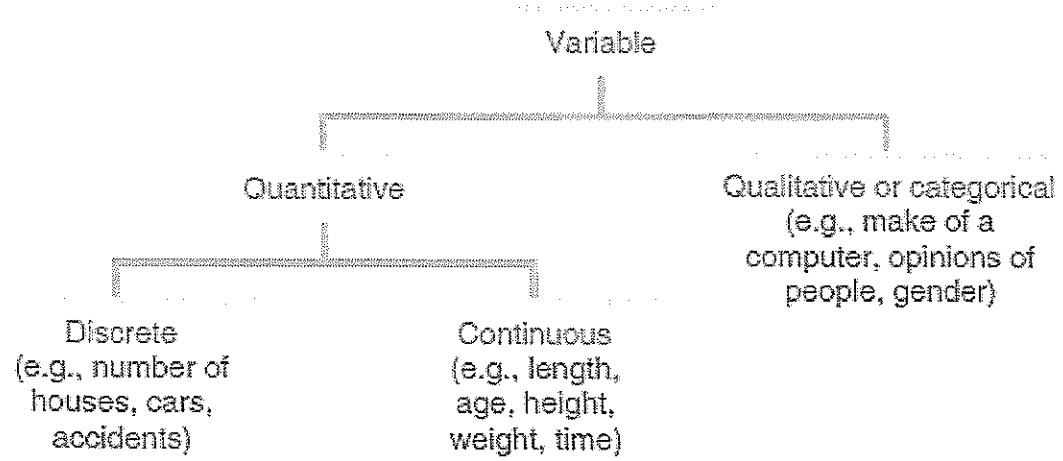


Figure 1.1 Types of variables.

EXERCISES

CONCEPTS AND PROCEDURES

1.7 Explain the meaning of the following terms.

a. Quantitative variable

b. Qualitative variable

c. Discrete variable

d. Continuous variable

e. Quantitative data

f. Qualitative data

APPLICATIONS

1.8 Indicate which of the following variables are quantitative and which are qualitative.

a. The amount of time a student spent studying for an exam

b. The amount of rain last year in 30 cities

c. The arrival status of an airline flight (early, on time, late, canceled) at an airport

d. A person's blood type

e. The amount of gasoline put into a car at a gas station

1.9 Classify the following quantitative variables as discrete or continuous.

a. The amount of time a student spent studying for an exam

b. The amount of rain last year in 30 cities

c. The amount of gasoline put into a car at a gas station

d. The number of customers in the line waiting for service at a bank at a given time