To: Probability and Statistics Students

Although summer is here, the math teachers at Notre Dame High School are already anticipating the next school year. We want to do everything possible to ensure that it will be very successful for you. It is important that the foundation for statistics is thoroughly understood. Therefore, the attached summer packet has been designed to introduce you to the course. It is required that you take the time to read the packet and complete all the exercises specified at the bottom of this cover letter. There is no need to print this packet out. Just hand-write (please do not type) your answers on a separate sheet of paper. **Your completed exercises will be collected the first day of school**, and will be graded as your first assignment. You will be tested on the material during the first week of classes.

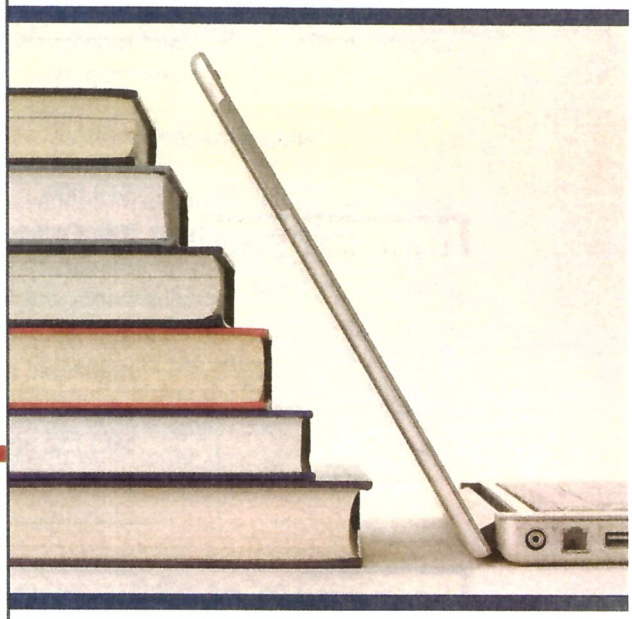
It is also required that you have your own graphing calculator that you bring to class on a daily basis. We have found that the TI-83+ and TI-84 calculator are the most helpful since they come with statistical packages as powerful as most of the computer programs used in college statistics courses.

We hope you have an enjoyable and relaxing summer. We are looking forward to seeing you during the next school year. Bring your completed summer packet with you on the first day of class so we can immediately begin a fun and challenging semester of Probability and Statistics!

Do the Review Exercises on pages 34 and 35 #1-58 all.

1

The Nature of Probability and Statistics



STATISTICS TODAY

Is Higher Education "Going Digital"?

Today many students take college courses online and use eBooks. Also, many students use a laptop, smartphone, or computer tablet in the classroom. With the increased use of technology, some questions about the effectiveness of this technology have been raised. For example,

How many colleges and universities offer online courses?

Do students feel that the online courses are equal in value to the traditional classroom presentations?

Approximately how many students take online courses now?

Will the number of students who take online courses increase in the future?

Has plagiarism increased since the advent of computers and the Internet?

Do laptops, smartphones, and tablets belong in the classroom?

Have colleges established any guidelines for the use of laptops, smartphones, and tablets?

To answer these questions, Pew Research Center conducted a study of college graduates and college presidents in 2011. The procedures they used and the results of the study are explained in this chapter. See Statistics Today—Revisited at the end of the chapter.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to

1. Demonstrate knowledge of statistical terms.
2. Differentiate between the two branches of statistics.
3. Identify types of data.
4. Identify the measurement level for each variable.
5. Identify the four basic sampling techniques.
6. Explain the difference between an observational and an experimental study.
7. Explain how statistics can be used and misused.
8. Explain the importance of computers and calculators in statistics.

Introduction

You may be familiar with probability and statistics through radio, television, newspapers, and magazines. For example, you may have read statements like the following found in newspapers.

- The FBI reported that violent crimes were down by 6.4% in 2011.
- *USA TODAY* reported that the average college graduate student loan debt was about $19,000.
- The College Stress and Mental Illness Poll reported that 85% of college and university students reported feeling stress daily; 77% reported stress from school work, and 74% experienced stress from grades.
- The *Occupational Outlook Handbook* reported that the median hourly wage for a registered nurse was $31.10 per hour.
- *Reader's Digest* reported that the average cost of using a plasma television is $0.1152 per hour.
- In 2013, the number of sales of smartphones is estimated to be 832.5 million units globally.

Unusual Stats

Of people in the United States, 14% said that they feel happiest in June, and 14% said that they feel happiest in December.

Statistics is used in almost all fields of human endeavor. In sports, for example, a statistician may keep records of the number of yards a running back gains during a football game, or the number of hits a baseball player gets in a season. In other areas, such as public health, an administrator might be concerned with the number of residents who contract a new strain of flu virus during a certain year. In education, a researcher might want to know if new methods of teaching are better than old ones. These are only a few examples of how statistics can be used in various occupations.

Interesting Fact

Every day in the United States about 120 golfers claim that they made a hole-in-one.

Furthermore, statistics is used to analyze the results of surveys and as a tool in scientific research to make decisions based on controlled experiments. Other uses of statistics include operations research, quality control, estimation, and prediction.

Statistics is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

Historical Note

A Scottish landowner and president of the Board of Agriculture, Sir John Sinclair introduced the word *statistics* into the English language in the 1798 publication of his book on a statistical account of Scotland. The word *statistics* is derived from the Latin word *status*, which is loosely defined as a statesman.

Students study statistics for several reasons:

1. Like professional people, you must be able to read and understand the various statistical studies performed in your fields. To have this understanding, you must be knowledgeable about the vocabulary, symbols, concepts, and statistical procedures used in these studies.
2. You may be called on to conduct research in your field, since statistical procedures are basic to research. To accomplish this, you must be able to design experiments; collect, organize, analyze, and summarize data; and possibly make reliable predictions or forecasts for future use. You must also be able to communicate the results of the study in your own words.
3. You can also use the knowledge gained from studying statistics to become better consumers and citizens. For example, you can make intelligent decisions about what products to purchase based on consumer studies, about government spending based on utilization studies, and so on.

These reasons can be considered some of the goals for studying statistics.

It is the purpose of this chapter to introduce the goals for studying statistics by answering questions such as the following:

What are the branches of statistics?

What are data?

How are samples selected?

1–1 Descriptive and Inferential Statistics

OBJECTIVE 1

Demonstrate knowledge of statistical terms.

To gain knowledge about seemingly haphazard situations, statisticians collect information for *variables,* which describe the situation.

A **variable** is a characteristic or attribute that can assume different values.

Data are the values (measurements or observations) that the variables can assume. Variables whose values are determined by chance are called **random variables.**

Suppose that an insurance company studies its records over the past several years and determines that, on average, 3 out of every 100 automobiles the company insured were involved in accidents during a 1-year period. Although there is no way to predict the specific automobiles that will be involved in an accident (random occurrence), the company can adjust its rates accordingly, since the company knows the general pattern over the long run. (That is, on average, 3% of the insured automobiles will be involved in an accident each year.)

A collection of data values forms a **data set.** Each value in the data set is called a **data value** or a **datum.**

In statistics it is important to distinguish between a sample and a population.

Historical Note

The 1880 Census had so many questions on it that it took 10 years to publish the results.

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.

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

When data are collected from every subject in the population, it is called a *census.*

For example, every 10 years the United States conducts a census. The primary purpose of this census is to determine the apportionment of the seats in the House of Representatives.

The first census was conducted in 1790 and was mandated by Article 1, Section 2 of the Constitution. As the United States grew, the scope of the census also grew. Today the Census limits questions to populations, housing, manufacturing, agriculture, and mortality. The Census is conducted by the Bureau of the Census, which is part of the Department of Commerce.

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. See Figure 1–1.

However, the information obtained from a statistical sample is said to be *biased* if the results from the sample of a population are radically different from the results of a census of the population. Also, a sample is said to be biased if it does not represent the population from which it has been selected. The techniques used to properly select a sample are explained in Section 1–3.

OBJECTIVE 2

Differentiate between the two branches of statistics.

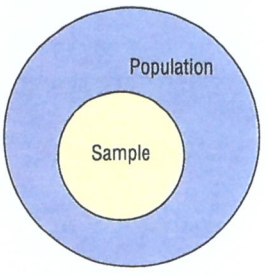
The body of knowledge called statistics is sometimes divided into two main areas, depending on how data are used. The two areas are

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.

FIGURE 1–1
Population and Sample



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.

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, Edmond Halley, published the first complete mortality tables. (Insurance companies use mortality tables to determine life insurance rates.)

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.

The 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 age 70 years of age. For this study, two groups of men over age 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 lung cancer in this group of volunteers was 10 times as great for smokers as for nonsmokers.

Unusual Stat

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

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.

EXAMPLE 1–1 Descriptive or Inferential Statistics

Determine whether descriptive or inferential statistics were used.

a. The average jackpot for the top five lottery winners was $367.6 million.

b. A study done by the American Academy of Neurology suggests that older people who had a high caloric diet more than doubled their risk of memory loss.

c. Based on a survey of 9317 consumers done by the National Retail Federation, the average amount that consumers spent on Valentine's Day in 2011 was $116.

d. Scientists at the University of Oxford in England found that a good laugh significantly raises a person's pain level tolerance.

SOLUTION

a. Descriptive statistics were used because this is an average, and it is based on data obtained from the top five lottery winners at this time.

b. Inferential statistics were used since this is a generalization made from a sample to a population.

c. Descriptive statistics were used since this is an average based on a sample of 9317 respondents.

d. Inferential statistics were used since an inference is made from a sample to a population.

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 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.

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

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 39 for the answers.

Exercises 1–1

1. Define statistics.

2. What is a variable?

3. What is meant by a census?

4. How does a population differ from a sample?

5. Explain the difference between descriptive and inferential statistics.

6. Name two areas where probability is used.

7. Why is information obtained from samples used more often than information obtained from populations?

8. What is meant by a biased sample?

For Exercises 9–17, determine whether descriptive or inferential statistics were used.

9. Because of the current economy, 49% of 18 to 34 year-olds have taken a job to pay the bills. (Source: Pew Research Center)

10. In 2025, the world population is predicted to be 8 billion people. (Source: United Nations)

11. In 2011, there were 34 deaths from the avian flu. (Source: World Health Organization)

12. Based on a sample of 2739 respondents, it is estimated that pet owners spent a total of 14 billion dollars on veterinarian care for their pets. (Source: American Pet Products Association, Pet Owners Survey)

13. In 2011, 79% of U.S. adults used the Internet. (Source: Pew Research Center)

14. In 2010, a total of 68,905 people died from diabetes. (Source: Centers for Disease Control and Prevention)

15. In an online survey of 500 Virginia Tech students between spring 2010 and spring 2011, 31% said that they had missed class because of alcohol consumption. (Source: Center for Applied Behavior Systems at Virginia Tech)

16. In 2008–2009, a total of 260,327 U.S. students were studying abroad. (Source: Institute of International Education)

17. Forty-four percent of the people in the United States have type O blood. (Source: American Red Cross)

Extending the Concepts

18. Find three statistical studies and explain whether they used descriptive or inferential statistics.

19. Find a gambling game and explain how probability was used to determine the outcome.

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 have 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 variables that can be counted or measured.

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.

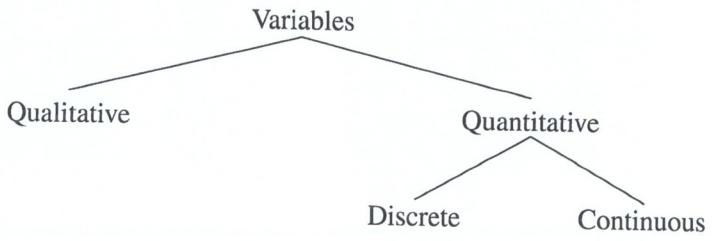
Quantitative variables can be further classified into two groups: discrete and continuous. *Discrete variables* can be assigned values such as 0, 1, 2, 3 and are said to be *countable*. Examples of discrete variables are the number of children in a family, the number of students in a classroom, and the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



EXAMPLE 1–2 Discrete or Continuous Variables

Classify each variable as a discrete variable or a continuous variable.

a. The highest wind speed of a hurricane
b. The weight of baggage on an airplane
c. The number of pages in a statistics book
d. The amount of money a person spends per year for online purchases

SOLUTION

a. Continuous, since wind speed must be measured
b. Continuous, since weight is measured
c. Discrete, since the number of pages is countable
d. Discrete, since the smallest value that money can assume is in cents

Unusual Stat

Fifty-two percent of Americans live within 50 miles of a coastal shoreline.

Since continuous data must be measured, answers must be rounded because of the limits of the measuring device. Usually, answers are rounded to the nearest given unit. For example, heights might be rounded to the nearest inch, weights to the nearest ounce, etc. Hence, a recorded height of 73 inches could mean any measure from 72.5 inches up to but not including 73.5 inches. Thus, the boundary of this measure is given as 72.5–73.5 inches. The **boundary** of a number, then, is defined as a class in which a data value would be placed before the data value was rounded. *Boundaries are written for convenience as 72.5–73.5 but are understood to mean all values up to but not including 73.5.* Actual data values of 73.5 would be rounded to 74 and would be included in a class with boundaries of 73.5 up to but not including 74.5, written as 73.5–74.5. As another example, if a recorded weight is 86 pounds, the exact boundaries are 85.5 up to but not including 86.5, written as 85.5–86.5 pounds. Table 1–1 helps to clarify this concept. The boundaries of a continuous variable are given in one additional decimal place and always end with the digit 5.

TABLE 1–1 Recorded Values and Boundaries

Variable	Recorded value	Boundaries
Length	15 centimeters (cm)	14.5–15.5 cm
Temperature	86 degrees Fahrenheit (°F)	85.5–86.5°F
Time	0.43 second (sec)	0.425–0.435 sec
Mass	1.6 grams (g)	1.55–1.65 g

EXAMPLE 1–3 Class Boundaries

Find the boundaries of each variable.

a. 8.4 quarts
b. 138 mmHg
c. 137.63 mg/dL

SOLUTION

a. 8.35–8.45 quarts
b. 137.5–138.5 mmHg
c. 137.625–137.635 mg/dL

In addition to being classified as qualitative or quantitative, variables can be classified by how they are categorized, counted, or measured. For example, can the data be organized into specific categories, such as area of residence (rural, suburban, or urban)? Can the data values be ranked, such as first place, second place, etc.? Or are the values obtained from measurement, such as heights, IQs, or temperature? This type of classification—i.e., how variables are categorized, counted, or measured—uses **measurement scales,** and four common types of scales are used: nominal, ordinal, interval, and ratio.

OBJECTIVE 4

Identify the measurement level for each variable.

The first level of measurement is called the *nominal level* of measurement. A sample of college instructors classified according to subject taught (e.g., English, history, psychology, or mathematics) is an example of nominal-level measurement. Classifying survey subjects as male or female is another example of nominal-level measurement. No ranking or order can be placed on the data. Classifying residents according to zip codes is also an example of the nominal level of measurement. Even though numbers are assigned as zip codes, there is no meaningful order or ranking. Other examples of nominal-level data are political party (Democratic, Republican, Independent, etc.), religion (Christianity, Judaism, Islam, etc.), and marital status (single, married, divorced, widowed, separated).

Unusual Stat

Sixty-three percent of us say we would rather hear the bad news first.

The **nominal level of measurement** classifies data into mutually exclusive (nonoverlapping) categories in which no order or ranking can be imposed on the data.

Historical Note

When data were first analyzed statistically by Karl Pearson and Francis Galton, almost all were continuous data. In 1899, Pearson began to analyze discrete data. Pearson found that some data, such as eye color, could not be measured, so he termed such data *nominal data.* Ordinal data were introduced by a German numerologist Frederich Mohs in 1822 when he introduced a hardness scale for minerals. For example, the hardest stone is the diamond, which he assigned a hardness value of 1500. Quartz was assigned a hardness value of 100. This does not mean that a diamond is 15 times harder than quartz. It only means that a diamond is harder than quartz. In 1947, a psychologist named Stanley Smith Stevens made a further division of continuous data into two categories, namely, interval and ratio.

The next level of measurement is called the *ordinal level.* Data measured at this level can be placed into categories, and these categories can be ordered, or ranked. For example, from student evaluations, guest speakers might be ranked as superior, average, or poor. Floats in a homecoming parade might be ranked as first place, second place, etc. *Note that precise measurement of differences in the ordinal level of measurement* does not *exist.* For instance, when people are classified according to their build (small, medium, or large), a large variation exists among the individuals in each class.

Other examples of ordinal data are letter grades (A, B, C, D, F).

The **ordinal level of measurement** classifies data into categories that can be ranked; however, precise differences between the ranks do not exist.

The third level of measurement is called the *interval level.* This level differs from the ordinal level in that precise differences do exist between units. For example, many standardized psychological tests yield values measured on an interval scale. IQ is an example of such a variable. There is a meaningful difference of 1 point between an IQ of 109 and an IQ of 110. Temperature is another example of interval measurement, since there is a meaningful difference of 1°F between each unit, such as 72 and 73°F. *One property is lacking in the interval scale: There is no true zero.* For example, IQ tests do not measure people who have no intelligence. For temperature, 0°F does not mean no heat at all.

The **interval level of measurement** ranks data, and precise differences between units of measure do exist; however, there is no meaningful zero.

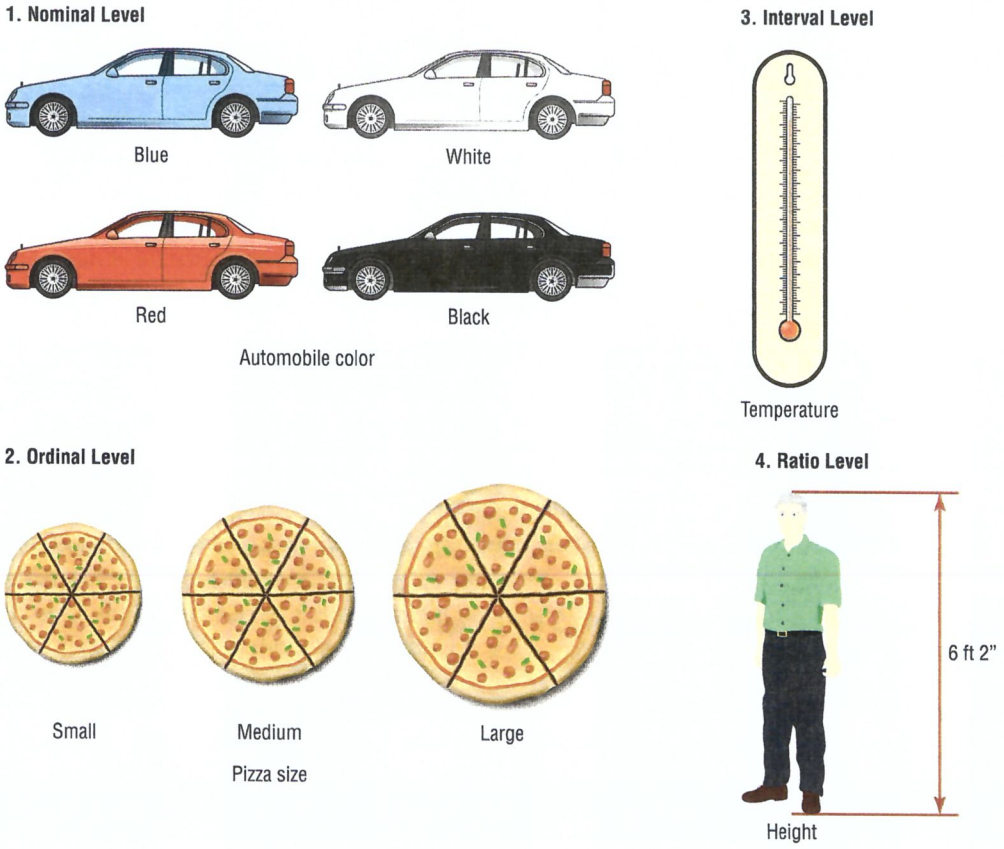
The final level of measurement is called the *ratio level.* Examples of ratio scales are those used to measure height, weight, area, and number of phone calls received. Ratio scales have differences between units (1 inch, 1 pound, etc.) and a true zero. In addition, the ratio scale contains a true ratio between values. For example, if one person can lift 200 pounds and another can lift 100 pounds, then the ratio between them is 2 to 1. Put another way, the first person can lift twice as much as the second person.

The **ratio level of measurement** possesses all the characteristics of interval measurement, and there exists a true zero. In addition, true ratios exist when the same variable is measured on two different members of the population.

TABLE 1–2 Examples of Measurement Scales

Nominal-level data	Ordinal-level data	Interval-level data	Ratio-level data
Zip code Gender (male, female) Eye color (blue, brown, green, hazel) Political affiliation Religious affiliation Major field (mathematics, computers, etc.) Nationality	Grade (A, B, C, D, F) Judging (first place, second place, etc.) Rating scale (poor, good, excellent) Ranking of tennis players	SAT score IQ Temperature	Height Weight Time Salary Age

FIGURE 1–2
Measurement Scales



There is not complete agreement among statisticians about the classification of data into one of the four categories. For example, some researchers classify IQ data as ratio data rather than interval. Also, data can be altered so that they fit into a different category. For instance, if the incomes of all professors of a college are classified into the three categories of low, average, and high, then a ratio variable becomes an ordinal variable. Table 1–2 gives some examples of each type of data. See Figure 1–2.

EXAMPLE 1–4 Measurement Levels

What level of measurement would be used to measure each variable?

a. The ages of patients in a local hospital
b. The ratings of movies released this month
c. Colors of athletic shirts sold by Oak Park Health Club
d. Temperatures of hot tubs in local health clubs

SOLUTION

a. Ratio
b. Ordinal
c. Nominal
d. Interval

Applying the Concepts 1–2

Fatal Transportation Injuries

Read the following information about the number of fatal accidents for the transportation industry in 2010, and answer each question.

Industry	Number of fatalities
Highway accidents	968
Railway accidents	44
Water vehicle accidents	52
Aircraft accidents	151

Source: Bureau of Labor Statistics.

1. Name the variables under study.
2. Categorize each variable as quantitative or qualitative.
3. Categorize each quantitative variable as discrete or continuous.
4. Identify the level of measurement for each variable.
5. The railroad has the fewest fatalities for 2010. Does that mean railroads have fewer accidents than the other industries?
6. What factors other than safety influence a person's choice of transportation?
7. From the information given, comment on the relationship between the variables.

See page 39 for the answers.

Exercises 1–2

1. Explain the difference between qualitative variables and quantitative variables.

2. Explain the difference between discrete and continuous variables.

3. Why are continuous variables rounded when they are used in statistical studies?

4. Name and define the four types of measurement levels used in statistics.

For Exercises 5–10, determine whether the data are qualitative or quantitative.

5. Color of football uniforms

6. Pizza sizes (small, medium, and large)

7. Cholesterol counts for individuals

8. Lightbulb wattage

9. Genders of movie stars

10. Ratings of teachers

For Exercises 11–16, determine whether the data are discrete or continuous.

11. Number of text messages sent in a month

12. Systolic blood pressure readings

13. Heights of basketball players on a local high school team

14. Number of phone calls received during one day at a poison control center

15. Number of students in the mathematics classes during the fall semester at your school for a particular school year

16. Temperatures at a seashore resort

For Exercises 17–22, give the boundaries of each value.

17. 7 feet

18. 6.3 millimeters

19. 143 miles

20. 3.25 pounds

21. 25.8 yards

22. 19 quarts

For Exercises 23–30, classify each as nominal-level, ordinal-level, interval-level, or ratio-level measurement.

23. Telephone numbers

24. Leap years: . . . 2008, 2012, 2016, . . .

25. Amounts of money spent on a medical checkup

26. Scores on a statistical final exam

27. Ratings of fiction books—excellent, good, fair, poor

28. Blood types—O, A, B, AB

29. Online spending

30. Number of flowers in a bouquet

1–3 Data Collection and Sampling Techniques

OBJECTIVE 5

Identify the four basic sampling techniques.

In research, statisticians use data in many different ways. As stated previously, data can be used to describe situations or events. For example, a manufacturer might want to know something about the consumers who will be purchasing his product so he can plan an effective marketing strategy. In another situation, the management of a company might survey its employees to assess their needs in order to negotiate a new contract with the employees' union. Data can be used to determine whether the educational goals of a school district are being met. Finally, trends in various areas, such as the stock market, can be analyzed, enabling prospective buyers to make more intelligent decisions concerning what stocks to purchase. These examples illustrate a few situations where collecting data will help people make better decisions on courses of action.

Data can be collected in a variety of ways. One of the most common methods is through the use of surveys. Surveys can be done by using a variety of methods. Three of the most common methods are the telephone survey, the mailed questionnaire, and the personal interview.

Historical Note

A pioneer in census taking was Pierre-Simon de Laplace. In 1780, he developed the Laplace method of estimating the population of a country. The principle behind his method was to take a census of a few selected communities and to determine the ratio of the population to the number of births in these communities. (Good birth records were kept.) This ratio would be used to multiply the number of births in the entire country to estimate the number of citizens in the country.

Telephone surveys have an advantage over personal interview surveys in that they are less costly. Also, people may be more candid in their opinions since there is no face-to-face contact. A major drawback to the telephone survey is that some people in the population will not have phones or will not answer when the calls are made; hence, not all people have a chance of being surveyed. Also, many people now have unlisted numbers and cell phones, so they cannot be surveyed. Finally, even the tone of voice of the interviewer might influence the response of the person who is being interviewed.

Mailed questionnaire surveys can be used to cover a wider geographic area than telephone surveys or personal interviews since mailed questionnaire surveys are less expensive to conduct. Also, respondents can remain anonymous if they desire. Disadvantages of mailed questionnaire surveys include a low number of responses and inappropriate answers to questions. Another drawback is that some people may have difficulty reading or understanding the questions.

Personal interview surveys have the advantage of obtaining in-depth responses to questions from the person being interviewed. One disadvantage is that interviewers must be trained in asking questions and recording responses, which makes the personal interview survey more costly than the other two survey methods. Another disadvantage is that the interviewer may be biased in his or her selection of respondents.

Data can also be collected in other ways, such as *surveying records* or *direct observation* of situations.

As stated in Section 1–1, researchers use samples to collect data and information about a particular variable from a large population. Using samples saves time and money and in some cases enables the researcher to get more detailed information about a particular subject. Remember, samples cannot be selected in haphazard ways because the information obtained might be biased. For example, interviewing people on a street corner during the day would not include responses from people working in offices at that time or from people attending school; hence, not all subjects in a particular population would have a chance of being selected.

To obtain samples that are unbiased—i.e., that give each subject in the population an equally likely chance of being selected—statisticians use four basic methods of sampling: random, systematic, stratified, and cluster sampling.

Historical Note

The first census in the United States was conducted in 1790. Its purpose was to ensure proper Congressional representation.

Random Sampling

A **random sample** is a sample in which all members of the population have an equal chance of being selected.

Random samples are selected by using chance methods or random numbers. One such method is to number each subject in the population. Then place numbered cards in a bowl, mix them thoroughly, and select as many cards as needed. The subjects whose numbers are selected constitute the sample. Since it is difficult to mix the cards thoroughly, there is a chance of obtaining a biased sample. For this reason, statisticians use another method of obtaining numbers. They generate random numbers with a computer or calculator. Before the invention of computers, random numbers were obtained from tables.

Some two-digit random numbers are shown in Table 1–3. To select a random sample of, say, 15 subjects out of 85 subjects, it is necessary to number each subject from 01 to 85. Then select a starting number by closing your eyes and placing your finger on a number in the table. (Although this may sound somewhat unusual, it enables us to find a starting number at random.) In this case suppose your finger landed on the number 12 in the second column. (It is the sixth number down from the top.) Then proceed downward until you have selected 15 different numbers between 01 and 85. When you reach the bottom of the column, go to the top of the next column. If you select a number greater than 85 or the number 00 or a duplicate number, just omit it. In our example, we will use the subjects numbered 12, 27, 75, 62, 57, 13, 31, 06, 16, 49, 46, 71, 53, 41, and 02. A more detailed procedure for selecting a random sample using a table of random numbers is given in Chapter 14, using Table D in Appendix A.

Systematic Sampling

A **systematic sample** is a sample obtained by selecting every k^{th} member of the population where k is a counting number.

Researchers obtain systematic samples by numbering each subject of the population and then selecting every k^{th} subject. For example, suppose there were 2000 subjects in the

SPEAKING OF STATISTICS

The Worst Day for Weight Loss

Many overweight people have difficulty losing weight. *Prevention* magazine reported that researchers from Washington University School of Medicine studied the diets of 48 adult weight loss participants. They used food diaries, exercise monitors, and weigh-ins. They found that the participants ate an average of 236 more calories on Saturdays than they did on the other weekdays. This would amount to a weight gain of 9 pounds per year. So if you are watching your diet, be careful on Saturdays.

Are the statistics reported in this study descriptive or inferential in nature? What type of variables are used here?

TABLE 1–3 Random Numbers

79	41	71	93	60	35	04	67	96	04	79	10	86
26	52	53	13	43	50	92	09	87	21	83	75	17
18	13	41	30	56	20	37	74	49	56	45	46	83
19	82	02	69	34	27	77	34	24	93	16	77	00
14	57	44	30	93	76	32	13	55	29	49	30	77
29	12	18	50	06	33	15	79	50	28	50	45	45
01	27	92	67	93	31	97	55	29	21	64	27	29
55	75	65	68	65	73	07	95	66	43	43	92	16
84	95	95	96	62	30	91	64	74	83	47	89	71
62	62	21	37	82	62	19	44	08	64	34	50	11
66	57	28	69	13	99	74	31	58	19	47	66	89
48	13	69	97	29	01	75	58	05	40	40	18	29
94	31	73	19	75	76	33	18	05	53	04	51	41
00	06	53	98	01	55	08	38	49	42	10	44	38
46	16	44	27	80	15	28	01	64	27	89	03	27
77	49	85	95	62	93	25	39	63	74	54	82	85
81	96	43	27	39	53	85	61	12	90	67	96	02
40	46	15	73	23	75	96	68	13	99	49	64	11

population and a sample of 50 subjects was needed. Since $2000 \div 50 = 40$, then $k = 40$, and every 40th subject would be selected; however, the first subject (numbered between 1 and 40) would be selected at random. Suppose subject 12 were the first subject selected; then the sample would consist of the subjects whose numbers were 12, 52, 92, etc., until 50 subjects were obtained. When using systematic sampling, you must be careful about how the subjects in the population are numbered. If subjects were arranged in a manner such as wife, husband, wife, husband, and every 40th subject were selected, the sample would consist of all husbands. Numbering is not always necessary. For example, a researcher may select every 10th item from an assembly line to test for defects.

Stratified Sampling

A **stratified sample** is a sample obtained by dividing the population into subgroups or strata according to some characteristic relevant to the study. (There can be several subgroups.) Then subjects are selected from each subgroup.

Samples within the strata should be randomly selected. For example, suppose the president of a two-year college wants to learn how students feel about a certain issue. Furthermore, the president wishes to see if the opinions of first-year students differ from those of second-year students. The president will randomly select students from each subgroup to use in the sample.

Cluster Sampling

A cluster sample is obtained by dividing the population into sections or clusters and then selecting one or more clusters and using all members in the cluster(s) as the members of the sample.

Historical Note

In 1936, the *Literary Digest,* on the basis of a biased sample of its subscribers, predicted that Alf Landon would defeat Franklin D. Roosevelt in the upcoming presidential election. Roosevelt won by a landslide. The magazine ceased publication the following year.

Here the population is divided into groups or clusters by some means such as geographic area or schools in a large school district. Then the researcher randomly selects some of these clusters and uses all members of the selected clusters as the subjects of the samples. Suppose a researcher wishes to survey apartment dwellers in a large city. If there are 10 apartment buildings in the city, the researcher can select at random 2 buildings from the 10 and interview all the residents of these buildings. Cluster sampling is used when the population is large or when it involves subjects residing in a large geographic area. For example, if one wanted to do a study involving the patients in the hospitals in New York City, it would be very costly and time-consuming to try to obtain a random sample of patients since they would be spread over a large area. Instead, a few hospitals could be selected at random, and the patients in these hospitals would be interviewed in a cluster. See Figure 1–3.

The four basic sampling methods are summarized in Table 1–4.

Other Sampling Methods

Interesting Facts

Older Americans are less likely to sacrifice happiness for a higher-paying job. According to one survey, 38% of those aged 18–29 said they would choose more money over happiness, while only 3% of those over age 65 would.

In addition to the four basic sampling methods, researchers use other methods to obtain samples. One such method is called a **convenience sample.** Here a researcher uses subjects who are convenient. For example, the researcher may interview subjects entering a local mall to determine the nature of their visit or perhaps what stores they will be patronizing. This sample is probably not representative of the general customers for several reasons. For one thing, it was probably taken at a specific time of day, so not all customers entering the mall have an equal chance of being selected since they were not there when the survey was being conducted. But convenience samples can be representative of the population. If the researcher investigates the characteristics of the population and determines that the sample is representative, then it can be used.

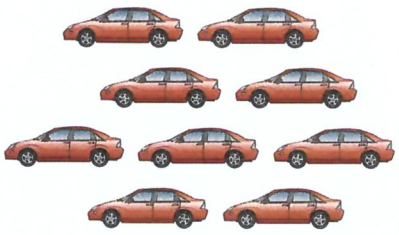
Another type of sample that is used in statistics is a *volunteer sample* or *self-selected sample.* Here respondents decide for themselves if they wish to be included in the sample. For example, a radio station in Pittsburgh asks a question about a situation and then asks people to call one number if they agree with the action taken or call another number if they disagree with the action. The results are then announced at the end of the day. Note that most often, only people with strong opinions will call. The station does explain that this is not a "scientific poll."

Since samples are not perfect representatives of the populations from which they are selected, there is always some error in the results. This error is called a *sampling error.*

Sampling error is the difference between the results obtained from a sample and the results obtained from the population from which the sample was selected.

FIGURE 1–3 Sampling Methods

1. Random

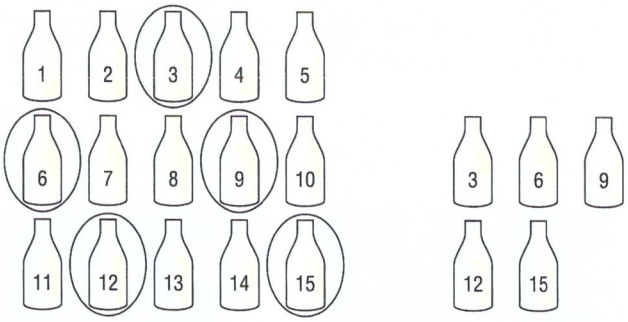


Population

Table D	Random Numbers			
10480	15011	01536	02011	81647
22368	46573	25595	85393	30995
24130	48360	22527	97265	76393
42167	93093	06243	61680	07856
37570	39975	81837	16656	06121
77921	06907	11008	42751	27750
99562	72905	56420	69994	98872
96301	91977	05463	07972	18876
89579	14342	63661	10281	17453
85475	36857	43342	53988	
28918	69578	88321		
63553	40961			

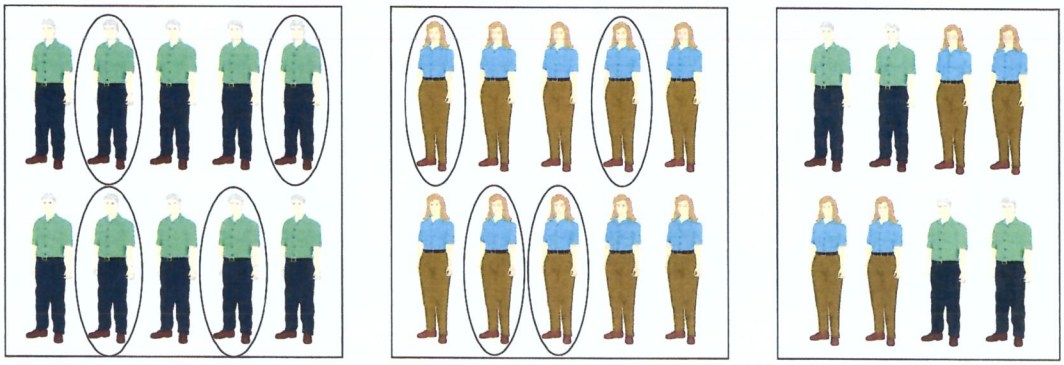
Sample

2. Systematic



Population Sample

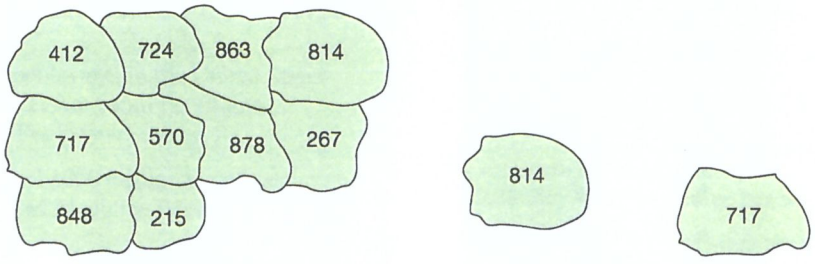
3. Stratified



Men Women Sample

Population

4. Cluster



Area codes Sample

Population

TABLE 1–4 Summary of Sampling Methods	
Random	Subjects are selected by random numbers.
Systematic	Subjects are selected by using every *k*th number after the first subject is randomly selected from 1 through *k*.
Stratified	Subjects are selected by dividing up the population into subgroups (strata), and subjects are randomly selected within subgroups.
Cluster	Subjects are selected by using an intact subgroup that is representative of the population.

For example, suppose you select a sample of full-time students at your college and find 56% are female. Then you go to the admissions office and get the genders of all full-time students that semester and find that 54% are female. The difference of 2% is said to be due to sampling error.

In most cases, this difference is unknown, but it can be estimated. This process is shown in Chapter 7.

There is another error that occurs in statistics called *nonsampling error.*

A **nonsampling error** occurs when the data are obtained erroneously or the sample is biased, i.e., nonrepresentative.

For example, data could be collected by using a defective scale. Each weight might be off by, say, 2 pounds. Also, recording errors can be made. Perhaps the researcher wrote an incorrect data value.

Caution and vigilance should be used when collecting data.

Other sampling techniques, such as *sequential sampling, double sampling,* and *multistage sampling,* are explained in Chapter 14, along with a more detailed explanation of the four basic sampling techniques.

EXAMPLE 1–5 Sampling Methods

State which sampling method was used.

a. Out of 10 hospitals in a municipality, a researcher selects one and collects records for a 24-hour period on the types of emergencies that were treated there.

b. A researcher divides a group of students according to gender, major field, and low, average, and high grade point average. Then she randomly selects six students from each group to answer questions in a survey.

c. The subscribers to a magazine are numbered. Then a sample of these people is selected using random numbers.

d. Every 10th bottle of Super-Duper Cola is selected, and the amount of liquid in the bottle is measured. The purpose is to see if the machines that fill the bottles are working properly.

SOLUTION

a. Cluster

b. Stratified

c. Random

d. Systematic

Applying the Concepts 1–3

American Culture and Drug Abuse

Assume you are a member of the Family Research Council and have become increasingly concerned about the drug use by professional sports players. You set up a plan and conduct a survey on how people believe the American culture (television, movies, magazines, and popular music) influences illegal drug use. Your survey consists of 2250 adults and adolescents from around the country. A consumer group petitions you for more information about your survey. Answer the following questions about your survey.

1. What type of survey did you use (phone, mail, or interview)?
2. What are the advantages and disadvantages of the surveying methods you did not use?
3. What type of scores did you use? Why?
4. Did you use a random method for deciding who would be in your sample?
5. Which of the methods (stratified, systematic, cluster, volunteer, or convenience) did you use?
6. Why was that method more appropriate for this type of data collection?
7. If a convenience sample were obtained consisting of only adolescents, how would the results of the study be affected?

See pages 39–40 for the answers.

Exercises 1–3

1. Name five ways that data can be collected.
2. What is meant by sampling error and nonsampling error?
3. Why are random numbers used in sampling, and how are random numbers generated?
4. Name and define the four basic sampling methods.

For Exercises 5–10, define a population that may have been used and explain how the sample might have been selected.

5. In 2011, 25% of people did not engage in regular physical activity. (Source: SQL Server Reporting Services)
6. The number one automobile that vehicle thieves prefer is the Honda Accord. (Source: LoJack Vehicle Theft Recovery Report, 2010)
7. There are 78 million pet owners in the United States who have dogs. (Source: American Pet Products Association National Pet Owners Survey)
8. Adults aged 19–50 need 1000 milligrams of calcium per day. (Source: Institute of Medicine Report)
9. Taking statins raises the risk of developing diabetes. (Source: *Journal of American Medical Association* and other sources)
10. The average January 2012 temperature in Boston was 34.2°F. This was 5.2° higher than the normal January average temperature. (Source: AccuWeather.com)

For Exercises 11–16, indentify the sampling method that was used.

11. To check the accuracy of a machine that is used for filling ice cream containers, every 20th bottle is selected and weighed.
12. To determine how long people exercise, a researcher interviews 5 people selected from a yoga class, 5 people selected from a weight-lifting class, 5 people selected from an aerobics class, and 5 people from swimming classes.
13. In a large school district, a researcher numbers all the full-time teachers and then randomly selects 30 teachers to be interviewed.
14. In a medical research study, a researcher selects a hospital and interviews all the patients that day.
15. Customers in the Sunrise Coffee Shop are asked how much they spend on coffee per week.
16. Ten counties in Pennsylvania are randomly selected to determine the average county real estate tax that the residents pay.

1–4 Experimental Design

OBJECTIVE 6

Explain the difference between an observational and an experimental study.

Observational and Experimental Studies

There are several different ways to classify statistical studies. This section explains two types of studies: *observational studies* and *experimental studies.*

In an **observational study,** the researcher merely observes what is happening or what has happened in the past and tries to draw conclusions based on these observations.

For example, data from the Motorcycle Industry Council (*USA TODAY*) stated that "Motorcycle owners are getting older and richer." Data were collected on the ages and incomes of motorcycle owners for the years 1980 and 1998 and then compared. The findings showed considerable differences in the ages and incomes of motorcycle owners for the two years. In this study, the researcher merely observed what had happened to the motorcycle owners over a period of time. There was no type of research intervention.

There are three main types of observational studies. When all the data are collected at one time, the study is called a *cross-sectional study*. When the data are collected using records obtained from the past, the study is called a *retrospective study*. Finally, if the data are collected over a period of time, say, past and present, the study is called a *longitudinal study*.

Observational studies have advantages and disadvantages. One advantage of an observational study is that it usually occurs in a natural setting. For example, researchers can observe people's driving patterns on streets and highways in large cities. Another advantage of an observational study is that it can be done in situations where it would be unethical or downright dangerous to conduct an experiment. Using observational studies, researchers can study suicides, rapes, murders, etc. In addition, observational studies can be done using variables that cannot be manipulated by the researcher, such as drug users versus nondrug users and right-handedness versus left-handedness.

Observational studies have disadvantages, too. As mentioned previously, since the variables are not controlled by the researcher, a definite cause-and-effect situation cannot be shown since other factors may have had an effect on the results. Observational studies can be expensive and time-consuming. For example, if one wanted to study the habitat of lions in Africa, one would need a lot of time and money, and there would be a certain amount of danger involved. Finally, since the researcher may not be using his or her own measurements, the results could be subject to the inaccuracies of those who collected the data. For example, if the researchers were doing a study of events that occurred in the 1800s, they would have to rely on information and records obtained by others from a previous era. There is no way to ensure the accuracy of these records.

The other type of study is called an *experimental study*.

In an **experimental study,** the researcher manipulates one of the variables and tries to determine how the manipulation influences other variables.

Interesting Fact

The safest day of the week for driving is Tuesday.

For example, a study conducted at Virginia Polytechnic Institute and presented in *Psychology Today* divided female undergraduate students into two groups and had the students perform as many sit-ups as possible in 90 sec. The first group was told only to "Do your best," while the second group was told to try to increase the actual number of sit-ups done each day by 10%. After four days, the subjects in the group who were given the vague instructions to "Do your best" averaged 43 sit-ups, while the group that was given the more specific instructions to increase the number of sit-ups by 10% averaged 56 sit-ups by the last day's session. The conclusion then was that athletes who were given specific goals performed better than those who were not given specific goals.

This study is an example of a statistical experiment since the researchers intervened in the study by manipulating one of the variables, namely, the type of instructions given to each group.

In a true experimental study, the subjects should be assigned to groups randomly. Also, the treatments should be assigned to the groups at random. In the sit-up study, the article did not mention whether the subjects were randomly assigned to the groups.

Sometimes when random assignment is not possible, researchers use intact groups. These types of studies are done quite often in education where already intact groups are available in the form of existing classrooms. When these groups are used, the study is said to be a **quasi-experimental study.** The treatments, though, should be assigned at random. Most articles do not state whether random assignment of subjects was used.

Statistical studies usually include one or more *independent variables* and one *dependent variable.*

The **independent variable** in an experimental study is the one that is being manipulated by the researcher. The independent variable is also called the **explanatory variable.** The resultant variable is called the **dependent variable** or the **outcome variable.**

The outcome variable is the variable that is studied to see if it has changed significantly because of the manipulation of the independent variable. For example, in the sit-up study, the researchers gave the groups two different types of instructions, general and specific. Hence, the independent variable is the type of instruction. The dependent variable, then, is the resultant variable, that is, the number of sit-ups each group was able to perform after four days of exercise. If the differences in the dependent or outcome variable are large and other factors are equal, these differences can be attributed to the manipulation of the independent variable. In this case, specific instructions were shown to increase athletic performance.

In the sit-up study, there were two groups. The group that received the special instruction is called the **treatment group** while the other is called the **control group.** The treatment group receives a specific treatment (in this case, instructions for improvement) while the control group does not.

Both types of statistical studies have advantages and disadvantages. Experimental studies have the advantage that the researcher can decide how to select subjects and how to assign them to specific groups. The researcher can also control or manipulate the independent variable. For example, in studies that require the subjects to consume a certain amount of medicine each day, the researcher can determine the precise dosages and, if necessary, vary the dosage for the groups.

Interesting Fact

The number of potholes in the United States is about 56 million.

There are several disadvantages to experimental studies. First, they may occur in unnatural settings, such as laboratories and special classrooms. This can lead to several problems. One such problem is that the results might not apply to the natural setting. The age-old question then is, "This mouthwash may kill 10,000 germs in a test tube, but how many germs will it kill in my mouth?"

Another disadvantage with an experimental study is the **Hawthorne effect.** This effect was discovered in 1924 in a study of workers at the Hawthorne plant of the Western Electric Company. In this study, researchers found that the subjects who knew they were participating in an experiment actually changed their behavior in ways that affected the results of the study.

Another problem when conducting statistical studies is called *confounding of variables* or *lurking variables.*

A **confounding variable** is one that influences the dependent or outcome variable but was not separated from the independent variable.

Researchers try to control most variables in a study, but this is not possible in some studies. For example, subjects who are put on an exercise program might also improve their diet unbeknownst to the researcher and perhaps improve their health in other ways not due to exercise alone. Then diet becomes a confounding variable.

When you read the results of statistical studies, decide if the study was observational or experimental. Then see if the conclusion follows logically, based on the nature of these studies.

Another factor that can influence statistical experiments is called the *placebo effect.* Here the subjects used in the study respond favorably or show improvement due to the fact that they had been selected for the study. They could also be reacting to clues given unintentionally by the researchers. For example, in a study on knee pain done at the Houston VA Medical Center, researchers divided 180 patients into three groups. Two groups had surgery to remove damaged cartilage while those in the third group had simulated surgery. After two years, an equal number of patients in each group reported that they felt better after the surgery. Those patients who had simulated surgery were said to be responding to what is called the placebo effect.

To minimize the placebo effect, researchers use what is called *blinding.* In blinding, the subjects do not know whether they are receiving an actual treatment or a placebo. Many times researchers use a sugar pill that looks like a real medical pill. Often *double blinding* is used. Here both the subjects and the researchers are not told which groups are given the placebos.

Unusual Stat

The chance that someone will attempt to burglarize your home in any given year is 1 in 20.

Researchers use *blocking* to minimize variability when they suspect that there might be a difference between two or more blocks. For example, in the sit-up study mentioned earlier, if we think that men and women would respond differently to "Do your best" versus "Increase by 10% every day," we would divide the subjects into two blocks (men, women) and then randomize which subjects in each block get the treatment.

When subjects are assigned to groups randomly, and the treatments are assigned randomly, the experiment is said to be a *completely randomized design.*"

Some experiments use what is called a *matched-pair design.* Here one subject is assigned to a treatment group, and another subject is assigned to a control group. But, before the assignment, subjects are paired according to certain characteristics. In earlier years, studies used identical twins, assigning one twin to one group and the other twin to another group. Subjects can be paired on any characteristics such as ages, heights, and weights.

Another way to validate studies is to use *replication.* Here the same experiment is done in another part of the country or in another laboratory. The same study could also be done using adults who are not going to college instead of using college students. Then the results of the second study are compared to the ones in the original study to see if they are the same.

No matter what type of study is conducted, two studies on the same subject sometimes have conflicting conclusions. Why might this occur? An article entitled "Bottom Line: Is It Good for You?" (*USA TODAY Weekend*) states that in the 1960s studies suggested that margarine was better for the heart than butter since margarine contains less saturated fat and users had lower cholesterol levels. In a 1980 study, researchers found that butter was better than margarine since margarine contained trans-fatty acids, which are worse for the heart than butter's saturated fat. Then in a 1998 study, researchers found that margarine was better for a person's health. Now, what is to be believed? Should one use butter or margarine?

The answer here is that you must take a closer look at these studies. Actually, it is not the choice between butter and margarine that counts, but the type of margarine used. In the 1980s, studies showed that solid margarine contains trans-fatty acids, and scientists believe that they are worse for the heart than butter's saturated fat. In the 1998 study, liquid margarine was used. It is very low in trans-fatty acids, and hence it is more healthful than butter because trans-fatty acids have been shown to raise cholesterol. Hence, the conclusion is that it is better to use liquid margarine than solid margarine or butter.

Before decisions based on research studies are made, it is important to get all the facts and examine them in light of the particular situation.

The purpose of a statistical study is to gain and process information obtained from the study in order to answer specific questions about the subject being investigated.

Statistical researchers use a specific procedure to do statistical studies to obtain valid results.

The general guidelines for this procedure are as follows:

1. Formulate the purpose of the study.
2. Identify the variables for the study.
3. Define the population.
4. Decide what sampling method you will use to collect the data.
5. Collect the data.
6. Summarize the data and perform any statistical calculations needed.
7. Interpret the results.

There is also a formal way to write up the study procedure and the results obtained. This information is available on the online resources under "Writing the Research Report."

EXAMPLE 1–6 Experimental Design

Researchers randomly assigned 10 people to each of three different groups. Group 1 was instructed to write an essay about the hassles in their lives. Group 2 was instructed to write an essay about circumstances that made them feel thankful. Group 3 was asked to write an essay about events that they felt neutral about. After the exercise, they were given a questionnaire on their outlook on life. The researchers found that those who wrote about circumstances that made them feel thankful had a more optimistic outlook on life. The conclusion is that focusing on the positive makes you more optimistic about life in general. Based on this study, answer the following questions.

a. Was this an observational or experimental study?
b. What is the independent variable?
c. What is the dependent variable?
d. What may be a confounding variable in this study?
e. What can you say about the sample size?
f. Do you agree with the conclusion? Explain your answer.

SOLUTION

a. This is an experimental study since the variables (types of essays written) were manipulated.
b. The independent variable was the type of essay the participants wrote.
c. The dependent variable was the score on the life outlook questionnaire.
d. Other factors, such as age, upbringing, and income, can affect the results; however, the random assignment of subjects is helpful in eliminating these factors. See the answer for the next question.
e. In this study, the sample uses 30 participants total.
f. Answers will vary.

OBJECTIVE 7

Explain how statistics can be used and misused.

Uses and Misuses of Statistics

As explained previously, statistical techniques can be used to describe data, compare two or more data sets, determine if a relationship exists between variables, test hypotheses, and make estimates about population characteristics. However, there is another aspect of statistics, and that is the misuse of statistical techniques to sell products that don't work properly, to attempt to prove something true that is really not true, or to get our attention by using statistics to evoke fear, shock, and outrage.

Two sayings that have been around for a long time illustrate this point:

"There are three types of lies—lies, damn lies, and statistics."

"Figures don't lie, but liars figure."

Just because we read or hear the results of a research study or an opinion poll in the media, this does not mean that these results are reliable or that they can be applied to any and all situations. For example, reporters sometimes leave out critical details such as the size of the sample used or how the research subjects were selected. Without this information, you cannot properly evaluate the research and properly interpret the conclusions of the study or survey.

It is the purpose of this section to show some ways that statistics can be misused. You should not infer that all research studies and surveys are suspect, but that there are many factors to consider when making decisions based on the results of research studies and surveys. Here are some ways that statistics can be misrepresented.

Suspect Samples The first thing to consider is the sample that was used in the research study. Sometimes researchers use very small samples to obtain information. Several years ago, advertisements contained such statements as "Three out of four doctors surveyed recommend brand such and such." If only 4 doctors were surveyed, the results could have been obtained by chance alone; however, if 100 doctors were surveyed, the results might be quite different.

Not only is it important to have a sample size that is large enough, but also it is necessary to see how the subjects in the sample were selected. As stated previously, studies using volunteers sometimes have a built-in bias. Volunteers generally do not represent the population at large. Sometimes they are recruited from a particular socioeconomic background, and sometimes unemployed people volunteer for research studies to get a stipend. Studies that require the subjects to spend several days or weeks in an environment other than their home or workplace automatically exclude people who are employed and cannot take time away from work. Sometimes only college students or retirees are used in studies. In the past, many studies have used only men, but have attempted to generalize the results to both men and women. Opinion polls that require a person to phone or mail in a response most often are not representative of the population in general, since only those with strong feelings for or against the issue usually call or respond by mail.

Another type of sample that may not be representative is the convenience sample. Educational studies sometimes use students in intact classrooms since it is convenient. Quite often, the students in these classrooms do not represent the student population of the entire school district.

When results are interpreted from studies using small samples, convenience samples, or volunteer samples, care should be used in generalizing the results to the entire population.

Ambiguous Averages In Chapter 3, you will learn that there are four commonly used measures that are loosely called *averages*. They are the *mean, median, mode,* and *midrange*. For the same data set, these averages can differ markedly. People who know this can, without lying, select the one measure of average that lends the most evidence to support their position.

Changing the Subject Another type of statistical distortion can occur when different values are used to represent the same data. For example, one political candidate who is running for reelection might say, "During my administration, expenditures increased a mere 3%." His opponent, who is trying to unseat him, might say, "During my opponent's administration, expenditures have increased a whopping $6,000,000." Here both figures are correct; however, expressing a 3% increase as $6,000,000 makes it sound

like a very large increase. Here again, ask yourself, Which measure better represents the data?

Detached Statistics A claim that uses a detached statistic is one in which no comparison is made. For example, you may hear a claim such as "Our brand of crackers has one-third fewer calories." Here, no comparison is made. One-third fewer calories than what? Another example is a claim that uses a detached statistic such as "Brand A aspirin works four times faster." Four times faster than what? When you see statements such as this, always ask yourself, Compared to what?

Implied Connections Many claims attempt to imply connections between variables that may not actually exist. For example, consider the following statement: "Eating fish may help to reduce your cholesterol." Notice the words *may help.* There is no guarantee that eating fish will definitely help you reduce your cholesterol.

"Studies suggest that using our exercise machine will reduce your weight." Here the word *suggest* is used; and again, there is no guarantee that you will lose weight by using the exercise machine advertised.

Another claim might say, "Taking calcium will lower blood pressure in some people." Note the word *some* is used. You may not be included in the group of "some" people. Be careful when you draw conclusions from claims that use words such as *may, in some people,* and *might help.*

Misleading Graphs Statistical graphs give a visual representation of data that enables viewers to analyze and interpret data more easily than by simply looking at numbers. In Chapter 2, you will see how some graphs are used to represent data. However, if graphs are drawn inappropriately, they can misrepresent the data and lead the reader to draw false conclusions. The misuse of graphs is also explained in Chapter 2.

Faulty Survey Questions When analyzing the results of a survey using questionnaires, you should be sure that the questions are properly written since the way questions are phrased can often influence the way people answer them. For example, the responses to a question such as "Do you feel that the North Huntingdon School District should build a new football stadium?" might be answered differently than a question such as "Do you favor increasing school taxes so that the North Huntingdon School District can build a new football stadium?" Each question asks something a little different, and the responses could be radically different. When you read and interpret the results obtained from questionnaire surveys, watch out for some of these common mistakes made in the writing of the survey questions.

In Chapter 14, you will find some common ways that survey questions could be misinterpreted by those responding, and could therefore result in incorrect conclusions.

In summary then, statistics, when used properly, can be beneficial in obtaining much information, but when used improperly, can lead to much misinformation. It is like your automobile. If you use your automobile to get to school or work or to go on a vacation, that's good. But if you use it to run over your neighbor's dog because it barks all night long and tears up your flower garden, that's not so good!

Applying the Concepts 1–4

Just a Pinch Between Your Cheek and Gum

As the evidence on the adverse effects of cigarette smoke grew, people tried many different ways to quit smoking. Some people tried chewing tobacco or, as it was called, smokeless tobacco. A small amount of tobacco was placed between the cheek and gum. Certain chemicals from the tobacco were absorbed into the bloodstream and gave the sensation of smoking cigarettes. This

prompted studies on the adverse effects of smokeless tobacco. One study in particular used 40 university students as subjects. Twenty were given smokeless tobacco to chew, and twenty given a substance that looked and tasted like smokeless tobacco, but did not contain any of the harmful substances. The students were randomly assigned to one of the groups. The students' blood pressure and heart rate were measured before they started chewing and 20 minutes after they had begun chewing. A significant increase in heart rate occurred in the group that chewed the smokeless tobacco. Answer the following questions.

1. What type of study was this (observational, quasi-experimental, or experimental)?
2. What are the independent and dependent variables?
3. Which was the treatment group?
4. Could the students' blood pressures be affected by knowing that they are part of a study?
5. List some possible confounding variables.
6. Do you think this is a good way to study the effect of smokeless tobacco?

See page 40 for the answers.

Exercises 1–4

1. Explain the difference between an observational and an experimental study.

2. Name and define the three types of observational studies.

3. List some advantages and disadvantages of an observational study.

4. List some advantages and disadvantages of an experimental study.

5. What is the difference between an experimental study and a quasi-experimental study?

6. What is the difference between independent variables and dependent variables?

7. Why are a treatment group and a control group used in a statistical study?

8. Explain the Hawthorne effect.

9. What is a confounding variable?

10. Define the placebo effect in a statistical study.

11. What is meant by blinding and double-blinding?

12. Why do researchers use randomization in statistical studies?

13. What is the difference between a completely randomized design and a matched-pair design?

14. Why is replication used in statistical studies?

For Exercises 15–18, determine whether an observational study or an experimental study was used.

15. Subjects are randomly assigned to three groups. Group 1 receives a placebo daily, group 2 receives a high dose of vitamin B daily, and group 3 receives a multivitamin daily. After 3 months, each participant is asked how many incidents of depression that she or he had, and the occurrences were compared.

16. In a large city, a researcher decides to check the police department records for 2 months to determine what types of vehicles were stolen.

17. In a public park, visitors were asked how many minutes a day they walked their dogs.

18. Subjects were randomly divided into two groups. One group was permitted 4 hours of sleep each night for one week. The other group was allowed to sleep at least 8 hours each night for one week. The blood pressures of both groups were taken at the end of the week and compared.

In Exercises 19–22, identify the independent variable and the dependent variable.

19. According to the *British Journal of Sports Medicine,* a regular 30-minute workout could slash your risk of catching a cold by 43%.

20. The *Journal of Behavioral Medicine* reported that sharing a hug and holding hands can limit the physical effects of stress such as soaring heart rate and elevated blood pressure.

21. A research study stated that meditation helps people make more-rational decisions.

22. A researcher found that the number of times people visited fast-food restaurants increased with increasing income. (Source: healthland.time.com)

For Exercises 23–26, suggest some confounding variables that the researcher might want to consider when doing a study.

23. *Psychology Today* magazine reports that the more intelligent a person is (based on IQ), the more willing the person is to make a cooperative choice rather than a selfish one.

24. *The New England Journal of Medicine* reported that when poor women move to better neighborhoods, they lower the risk of developing obesity and diabetes.

25. A researcher from the University of California found that gratitude can produce positive emotions that can lead to a longer, healthier life.

26. York University in Toronto, Canada, stated that people who had suffered from fibromyalgia were able to reduce their pain by participating in twice-weekly yoga sessions.

For Exercises 27–31, give a reason why the statement made might be misleading.

27. Our product will give you the perfect body.

28. Here is the whole truth about back pain.

29. Our pain medicine will give you 24 hours of pain relief.

30. By reading this book, you will increase your IQ by 20 points.

31. Eating 21 grams of fiber may help you to lose weight.

32. List the steps you should perform when conducting a statistical study.

33. **Beneficial Bacteria** According to a pilot study of 20 people conducted at the University of Minnesota, daily doses of a compound called arabinogalactan over a period of 6 months resulted in a significant increase in the beneficial lactobacillus species of bacteria. Why can't it be concluded that the compound is beneficial for the majority of people?

34. Comment on the following statement, taken from a magazine advertisement: "In a recent clinical study, Brand ABC (actual brand will not be named) was proved to be 1950% better than creatine!"

35. In an ad for women, the following statement was made: "For every 100 women, 91 have taken the road less traveled." Comment on this statement.

36. In many ads for weight loss products, under the product claims and in small print, the following statement is made: "These results are not typical." What does this say about the product being advertised?

37. In an ad for moisturizing lotion, the following claim is made: ". . . it's the number 1 dermatologist-recommended brand." What is misleading about this claim?

38. An ad for an exercise product stated: "Using this product will burn 74% more calories." What is misleading about this statement?

39. "Vitamin E is a proven antioxidant and may help in fighting cancer and heart disease." Is there anything ambiguous about this claim? Explain.

40. "Just 1 capsule of Brand X can provide 24 hours of acid control." (Actual brand will not be named.) What needs to be more clearly defined in this statement?

41. ". . . Male children born to women who smoke during pregnancy run a risk of violent and criminal behavior that lasts well into adulthood." Can we infer that smoking during pregnancy is responsible for criminal behavior in people?

42. **Caffeine and Health** In the 1980s, a study linked coffee to a higher risk of heart disease and pancreatic cancer. In the early 1990s, studies showed that drinking coffee posed minimal health threats. However, in 1994, a study showed that pregnant women who drank 3 or more cups of tea daily may be at risk for spontaneous abortion. In 1998, a study claimed that women who drank more than a half-cup of caffeinated tea every day may actually increase their fertility. In 1998, a study showed that over a lifetime, a few extra cups of coffee a day can raise blood pressure, heart rate, and stress (Source: "Bottom Line: Is It Good for You? Or Bad?" by Monika Guttman, *USA TODAY Weekend*). Suggest some reasons why these studies appear to be conflicting.

Extending the Concepts

43. Find an article that describes a statistical study, and identify the study as observational or experimental.

44. For the article that you used in Exercise 43, identify the independent variable(s) and dependent variable for the study.

45. For the article that you selected in Exercise 43, suggest some confounding variables that may have an effect on the results of the study.

46. Select a newspaper or magazine article that involves a statistical study, and write a paper answering these questions.

a. Is this study descriptive or inferential? Explain your answer.

b. What are the variables used in the study? In your opinion, what level of measurement was used to obtain the data from the variables?

c. Does the article define the population? If so, how is it defined? If not, how could it be defined?

d. Does the article state the sample size and how the sample was obtained? If so, determine the size of the sample and explain how it was selected. If not, suggest a way it could have been obtained.

e. Explain *in your own words* what procedure (survey, comparison of groups, etc.) might have been used to determine the study's conclusions.

f. Do you agree or disagree with the conclusions? State your reasons.

1–5 Computers and Calculators

OBJECTIVE 8

Explain the importance of computers and calculators in statistics.

In the past, statistical calculations were done with pencil and paper. However, with the advent of calculators, numerical computations became much easier. Computers do all the numerical calculation. All one does is to enter the data into the computer and use the appropriate command; the computer will print the answer or display it on the screen. Now the TI-84 Plus graphing calculator accomplishes the same thing.

There are many statistical packages available. This book uses Microsoft Excel and MINITAB. Instructions for using the TI-84 Plus graphing calculator, Excel, and MINITAB have been placed at the end of each relevant section, in subsections entitled Technology Step by Step.

You should realize that the computer and calculator merely give numerical answers and save the time and effort of doing calculations by hand. You are still responsible for understanding and interpreting each statistical concept. In addition, you should realize that the results come from the data and do not appear magically on the computer. Doing calculations by using the procedure tables will help you reinforce this idea.

The author has left it up to instructors to choose how much technology they will incorporate into the course.

Technology Step by Step

TI-84 Plus

Step by Step

The TI-84 Plus graphing calculator can be used for a variety of statistical graphs and tests.

General Information

To turn calculator on:
Press **ON** key.
To turn calculator off:
Press **2nd [OFF].**

To reset defaults only:

1. Press **2nd,** then **[MEM].**
2. Select **7,** then **2,** then **2.**

Optional. To reset settings on calculator and clear memory (*note:* this will clear all settings and programs in the calculator's memory):
Press **2nd,** then **[MEM].** Then press **7,** then **1,** then **2.**
(Also, the contrast may need to be adjusted after this.)
To adjust contrast (if necessary):
Press **2nd.** Then press and hold ▲ to darken or ▼ to lighten contrast.

To clear screen:
Press **CLEAR.**
(*Note:* This will return you to the screen you were using.)
To display a menu:
Press appropriate menu key. Example: **STAT.**
To return to home screen:
Press **2nd,** then **[QUIT].**
To move around on the screens:
Use the arrow keys.
To select items on the menu:
Press the corresponding number or move the cursor to the item, using the arrow keys. Then press **ENTER.**
(*Note:* In some cases, you do not have to press **ENTER,** and in other cases you may need to press **ENTER** twice.)

Entering Data

To enter single-variable data (clear the old list if necessary, see "Editing Data"):

1. Press **STAT** to display the Edit menu.
2. Press **ENTER** to select 1:Edit.
3. Enter the data in L_1 and press **ENTER** after each value.
4. After all data values are entered, press **STAT** to get back to the Edit menu or **2nd [QUIT]** to end.

Example TI1–1

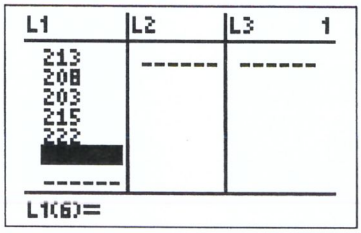
Enter the following data values in L_1: **213, 208, 203, 215, 222.**

To enter multiple-variable data:

The TI-84 Plus will take up to six lists designated L_1, L_2, L_3, L_4, L_5, and L_6.

1. To enter more than one set of data values, complete the preceding steps. Then move the cursor to L_2 by pressing the ▶ key.
2. Repeat the steps in the preceding part.

Output



Editing Data

To correct a data value before pressing **ENTER,** use ◀ and retype the value and press **ENTER.**
To correct a data value in a list after pressing **ENTER,** move the cursor to the incorrect value in list and type in the correct value. Then press **ENTER.**
To delete a data value in a list:
Move the cursor to a value and press **DEL.**
To insert a data value in a list:

1. Move cursor to position where data value is to be inserted; then press **2nd [INS].**
2. Type data value; then press **ENTER.**

To clear a list:

1. Press **STAT,** then **4.**
2. Enter list to be cleared. Example: To clear L_1, press **2nd [L_1].** Then press **ENTER.**

   (*Note:* To clear several lists, follow Step 1, but enter each list to be cleared, separating them with commas. To clear all lists at once, follow Step 1; then press **ENTER.**)

Sorting Data

To sort the data in a list:

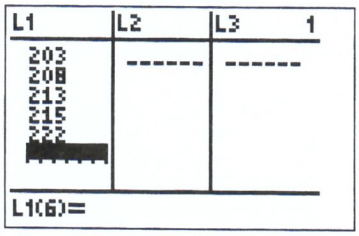
1. Enter the data in L_1.
2. Press **STAT 2** to get SortA to sort the list in ascending order.
3. Then press **2nd [L_1] ENTER.**

The calculator will display Done.

4. Press **STAT ENTER** to display the sorted list.

   (*Note:* The SortD or **3** sorts the list in descending order.)

Output



Example TI1–2

Sort in ascending order the data values entered in Example TI1–1.

EXCEL

Step by Step

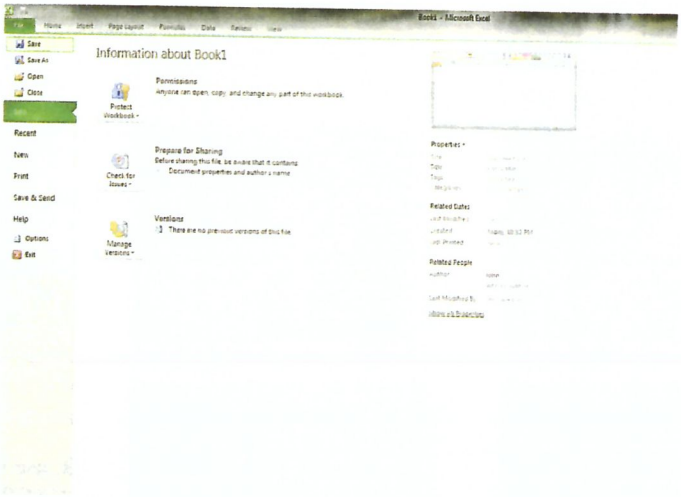
General Information

Microsoft Excel 2010 has two different ways to solve statistical problems. First, there are built-in functions, such as STDEV and CHITEST, available from the standard toolbar by clicking Formulas, and then selecting the Insert Function icon f_x. Another feature of Excel that is useful for calculating multiple statistical measures and performing statistical tests for a set of data is the Data Analysis command found in the Analysis ToolPak Add-in.

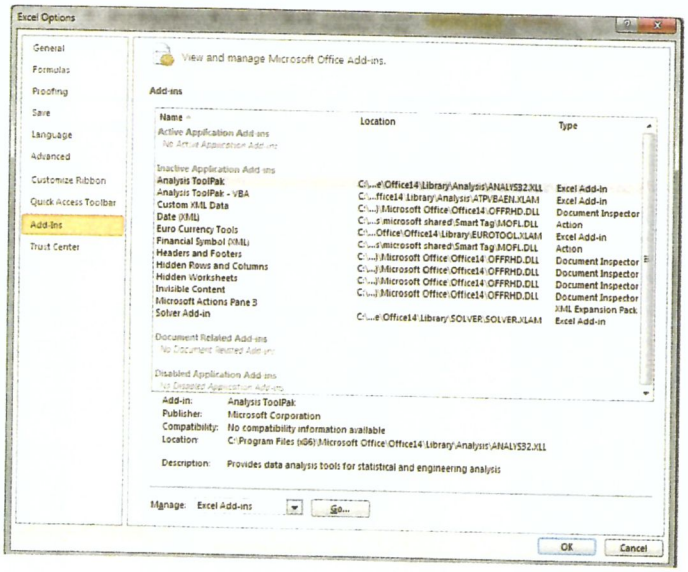
To load the Analysis ToolPak:

Excel's Analysis ToolPak Add-In

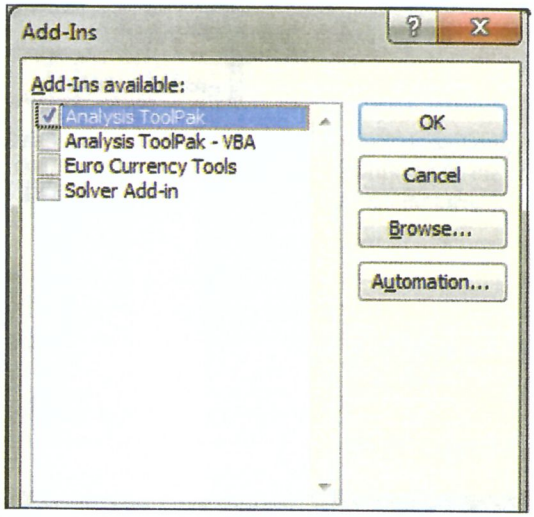
Click the File tab in the upper left-hand corner of an Excel workbook, then select Options in the left-hand panel.



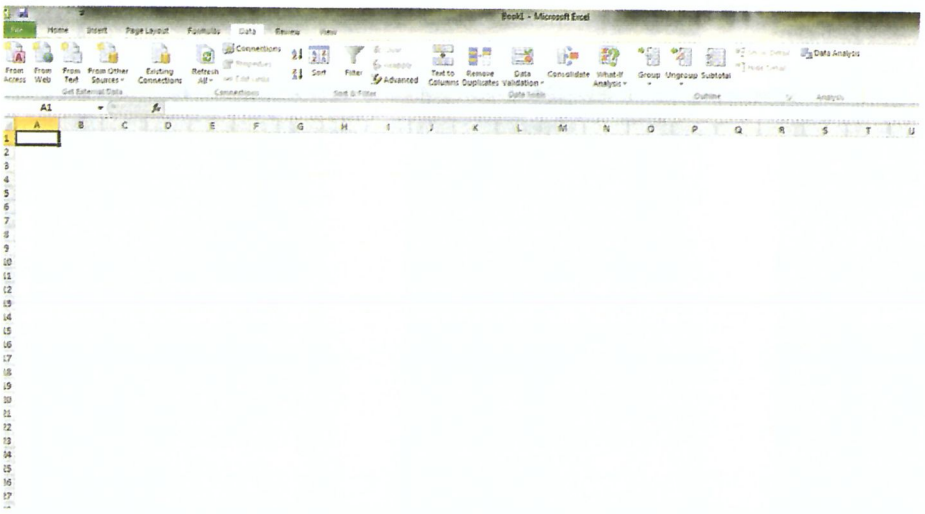
1. Click Add-Ins, and then click the Go button at the bottom of the Excel Options page to the right of the Manage tool.



2. Check the Analysis ToolPak Add-in and click OK.



3. Click the Data Tab. The Analysis ToolPak will appear in the Analysis group.

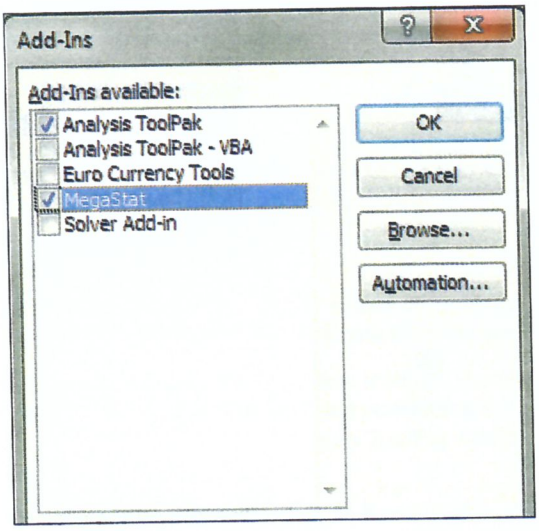


MegaStat

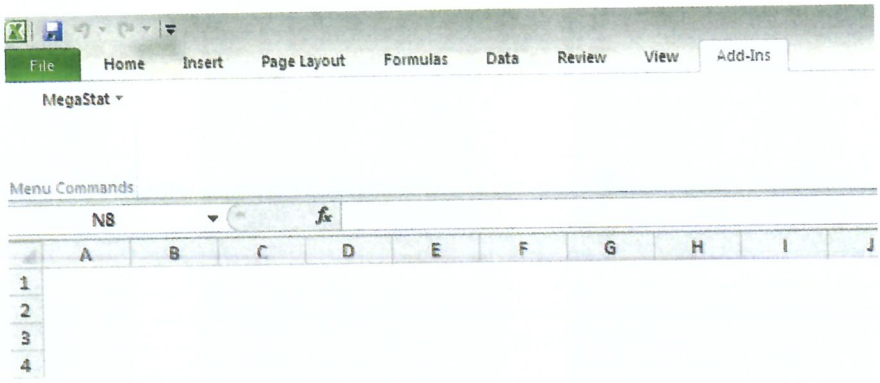
Later in this text you will encounter a few Excel Technology Step by Step operations that will require the use of the MegaStat Add-in for Excel. MegaStat can be purchased from www.mhhe.com/megastat.

1. Save the Zip file containing the MegaStat Excel Add-in file (MegaStat.xls) and the associated help file on your computer's hard drive.
2. Open the Excel software.
3. Click the File tab and select Options (as before with the installation of the Analysis ToolPak).
4. Click the Add-Ins button. MegaStat will not appear as an Application until first installation.
5. Click Go button next to the Manage (Add-Ins) Tool at the bottom of the Excel Options window.
6. Once the Add-Ins Checkbox appears, click Browse to locate the MegaStat.xla file on your computer's hard drive.

7. Select the MegaStat.xla file from the hard drive; click the Checkbox next to it and click OK.



8. The MegaStat Add-in will appear when you select the Add-Ins tab in the Toolbar.



Entering Data

1. Select a cell at the top of a column on an Excel worksheet where you want to enter data. When working with data values for a single variable, you will usually want to enter the values into a single column.
2. Type each data value and press **[Enter]** or **[Tab]** on your keyboard.

You can also add more worksheets to an Excel workbook by clicking the Insert Worksheet icon located at the bottom of an open workbook.

Example XL1–1: Opening an existing Excel workbook/worksheet

1. Open the Microsoft Office Excel 2010 program.
2. Click the File tab, then click Open.
3. Click the name of the library that contains the file, such as My documents, and then click Open.
4. Click the name of the Excel file that you want to open and then click Open.

Note: Excel files have the extension .xls

IINITAB
ep by Step

General Information

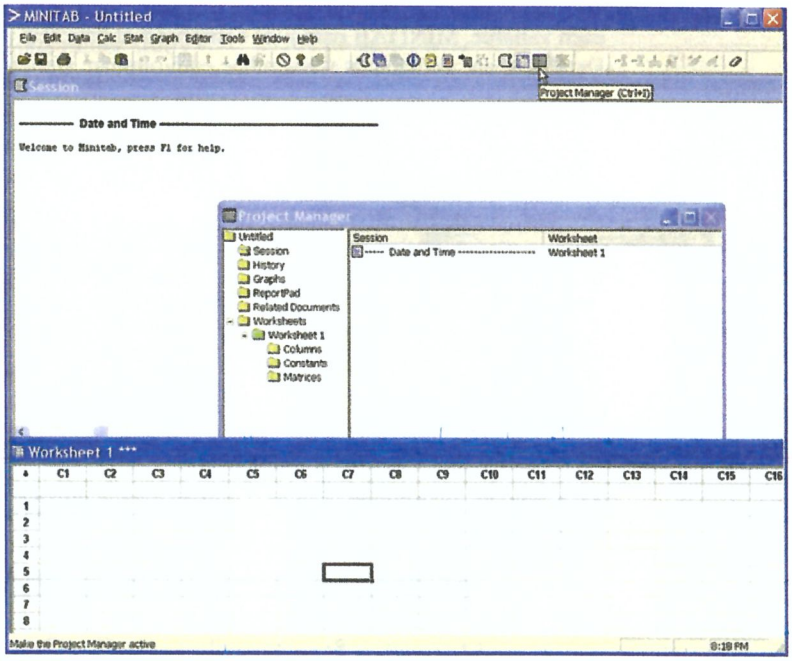
MINITAB statistical software provides a wide range of statistical analysis and graphing capabilities.

Take Note

In this text you will see captured MINITAB images from Windows computers running MINITAB Release 16. If you are using an earlier or later release of MINITAB, the screens you see on your computer may bear slight visual differences from the screens pictured in this text.

Start the Program

1. Click the Windows Start Menu, then All Programs.
2. Click the MINITAB folder and then click Minitab 16, the program icon. The program screen will look similar to the one shown here. You will see the Session Window, the Worksheet Window, and perhaps the Project Manager Window.
3. Click the Project Manager icon on the toolbar to bring the project manager to the front.



To use the program, data must be entered from the keyboard or from a file.

Entering Data from the Keyboard

In MINITAB, all the data for one variable are stored in a column. Step by step instructions for entering these data follow.

Data

213 208 203 215 222

1. Click in row 1 of Worksheet 1***. This makes the worksheet the active window and puts the cursor in the first cell. The small data entry arrow in the upper left-hand corner of the worksheet should be pointing down. If it is not, click it to change the direction in which the cursor will move when you press the [Enter] key.
2. Type in each number, pressing [Enter] after each entry, including the last number typed.
3. *Optional:* Click in the space above row 1 to type in **Weight,** the column label.

Worksheet 1 ***

	C1
	Weight
1	213
2	208
3	203
4	215
5	222

Save a Worksheet File

4. Click on the **File Menu.** *Note:* This is *not* the same as clicking the disk icon .
5. Click **Save Current Worksheet As . . .**
6. In the dialog box you will need to verify three items:
   a) Save in: Click on or type in the disk drive and directory where you will store your data. This may be a thumb drive such as E:\ or a hard-drive folder such as C:\MinitabData.
   b) File Name: Type in the name of the file, such as **MyData.**
   c) Save as Type: If using the Student Edition, we would need to use the portable file format. The default here is MINITAB. An extension of mtw is added to the name.

   Click [Save]. The name of the worksheet will change from Worksheet 1*** to MyData.MTP***. The triple asterisks indicate the active worksheet.

Open the Databank File

The raw data are shown in Appendix B. There is a row for each person's data and a column for each variable. MINITAB data files comprised of data sets used in this book, including the Databank, are available at www.mhhe.com/bluman. Here is how to get the data from a file into a worksheet.

1. Click **File>Open Worksheet.** A sequence of menu instructions will be shown this way.

   Note: This is *not* the same as clicking the file icon . If the dialog box says Open Project instead of Open Worksheet, click [Cancel] and use the correct menu item. The Open Worksheet dialog box will be displayed.
2. You must check three items in this dialog box.
   a) The Look In: dialog box should show the directory where the file is located.
   b) Make sure the Files of Type: shows the correct type, MINITAB [*.mtp].
   c) Double-click the file name in the list box Databank.mtp. A dialog box may inform you that a copy of this file is about to be added to the project. Click on the checkbox if you do not want to see this warning again.
3. Click the [OK] button. The data will be copied into a second worksheet. Part of the worksheet is shown here.

Databank.MTW ***

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11-T	C12-T
	ID-NUMBER	AGE	ED-LEVEL	SMOKING	EXERCISE	WEIGHT	SERUM-	SYSTOLIC	IQ	SODIUM	GENDER	MARITAL-ST
1	1	27	2	1	1	120	193	126	118	136	F	M
2	2	18	1	0	1	145	210	120	105	137	M	S
3	3	32	2	0	0	118	196	128	115	135	F	M

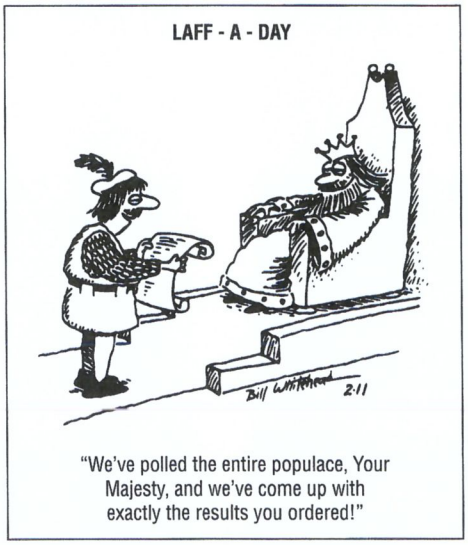
   a) You may maximize the window and scroll if desired.
   b) C12-T Marital Status has a T appended to the label to indicate alphanumeric data. MyData.MTW is not erased or overwritten. Multiple worksheets can be available; however, only the active worksheet is available for analysis.
4. To switch between the worksheets, select **Window>MyData.MTW.**
5. Select **File>Exit** to quit. To save the project, click [Yes].
6. Type in the name of the file, **Chapter01.** The Data Window, the Session Window, and settings are all in one file called a project. Projects have an extension of mpj instead of mtw.

   Clicking the disk icon on the menu bar is the same as selecting **File>Save Project.**

   Clicking the file icon is the same as selecting **File>Open Project.**
7. Click [Save]. The mpj extension will be added to the name. The computer will return to the Windows desktop. The two worksheets, the Session Window results, and settings are saved in this project file. When a project file is opened, the program will start up right where you left off.

Summary

- The two major areas of statistics are descriptive and inferential. Descriptive statistics includes the collection, organization, summarization, and presentation of data. Inferential statistics includes making inferences from samples to populations, estimations and hypothesis testing, determining relationships, and making predictions. Inferential statistics is based on *probability theory.* (1–1)*
- Data can be classified as qualitative or quantitative. Quantitative data can be either discrete or continuous, depending on the values they can assume. Data can also be measured by various scales. The four basic levels of measurement are nominal, ordinal, interval, and ratio. (1–2)
- Since in most cases the populations under study are large, statisticians use subgroups called samples to get the necessary data for their studies. There are four basic methods used to obtain samples: random, systematic, stratified, and cluster. (1–3)
- There are two basic types of statistical studies: observational studies and experimental studies. When conducting observational studies, researchers observe what is happening or what has happened and then draw conclusions based on these observations. They do not attempt to manipulate the variables in any way. (1–4)
- When conducting an experimental study, researchers manipulate one or more of the independent or explanatory variables and see how this manipulation influences the dependent or outcome variable. (1–4)
- Finally, the applications of statistics are many and varied. People encounter them in everyday life, such as in reading newspapers or magazines, listening to an MP3 player, or watching television. Since statistics is used in almost every field of endeavor, the educated individual should be knowledgeable about the vocabulary, concepts, and procedures of statistics. Also, everyone should be aware that statistics can be misused. (1–4)
- Today, computers and calculators are used extensively in statistics to facilitate the computations. (1–5)



"We've polled the entire populace, Your Majesty, and we've come up with exactly the results you ordered!"

*The numbers in parentheses indicate the chapter section where the material is explained.

Important Terms

blinding 20
blocks 20
boundary 7
census 3
cluster sample 14
completely randomized design 20
confounding variable 19
continuous variables 6
control group 19
convenience sample 14
cross-sectional study 18
data 3
data set 3
data value or datum 3
dependent variable 19
descriptive statistics 3
discrete variables 6
double blinding 20
experimental study 18
explanatory variable 19
Hawthorne effect 19
hypothesis testing 4
independent variable 19
inferential statistics 4
interval level of measurement 8
longitudinal study 18
matched-pair design 20
measurement scales 8
nominal level of measurement 8

nonsampling error 16
observational study 18
ordinal level of measurement 8
outcome variable 19
placebo effect 20
population 3
probability 4
qualitative variables 6
quantitative variables 6
quasi-experimental study 19
random sample 12
random variable 3
ratio level of measurement 8
replication 20
retrospective study 18
sample 3
sampling error 14
statistics 2
stratified sample 13
systematic sample 12
treatment group 19
variable 3
volunteer sample 14

Review Exercises

Section 1–1

For Exercises 1–8, state whether descriptive or inferential statistics has been used.

1. By 2040 at least 3.5 billion people will run short of water (World Future Society).
2. In a sample of 100 on-the-job fatalities, 90% of the victims were men.
3. In a survey of 1000 adults, 34% said that they posted notes on social media websites (Source: AARP Survey).
4. In a poll of 3036 adults, 32% said that they got a flu shot at a retail clinic (Source: Harris Interactive Poll).
5. Allergy therapy makes bees go away (Source: *Prevention*).
6. Drinking decaffeinated coffee can raise cholesterol levels by 7% (Source: American Heart Association).
7. The average stay in a hospital for 2000 patients who had circulatory system problems was 4.7 days.
8. Experts say that mortgage rates may soon hit bottom (Source: *USA TODAY*).

Section 1–2

For Exercises 9–18, classify each as nominal-level, ordinal-level, interval-level, or ratio-level measurement.

9. Pages in the 25 best-selling mystery novels.
10. Rankings of golfers in a tournament.
11. Temperatures inside 10 pizza ovens.
12. Weights of selected cell phones.
13. Salaries of the coaches in the NFL.
14. Times required to complete a chess game.
15. Ratings of textbooks (poor, fair, good, excellent).
16. Number of amps delivered by battery chargers.
17. Ages of children in a day care center.
18. Categories of magazines in a physician's office (sports, women's, health, men's, news).

For Exercises 19–26, classify each variable as qualitative or quantitative.

19. Marital status of nurses in a hospital.
20. Time it takes to run a marathon.
21. Weights of lobsters in a tank in a restaurant.
22. Colors of automobiles in a shopping center parking lot.
23. Ounces of ice cream in a large milkshake.
24. Capacity of the NFL football stadiums.
25. Ages of people living in a personal care home.
26. Different vitamins taken.

For Exercises 27–34, classify each variable as discrete or continuous.

27. Number of pizzas sold by Pizza Express each day.
28. Relative humidity levels in operating rooms at local hospitals.
29. Number of bananas in a bunch at several local supermarkets.
30. Lifetimes (in hours) of 15 iPod batteries.
31. Weights of the backpacks of first-graders on a school bus.
32. Number of students each day who make appointments with a math tutor at a local college.
33. Blood pressures of runners in a marathon.
34. Ages of children in a preschool.

For Exercises 35–38, give the boundaries of each value.

35. 36 inches.
36. 105.4 miles.
37. 72.6 tons.
38. 5.27 centimeters.

Section 1–3

For Exercises 39–44, classify each sample as random, systematic, stratified, cluster, or other.

39. In a large school district, all teachers from two buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years.

40. Every seventh customer entering a shopping mall is asked to select her or his favorite store.

41. Nursing supervisors are selected using random numbers to determine annual salaries.

42. Every 100th hamburger manufactured is checked to determine its fat content.

43. Mail carriers of a large city are divided into four groups according to gender (male or female) and according to whether they walk or ride on their routes. Then 10 are selected from each group and interviewed to determine whether they have been bitten by a dog in the last year.

44. People are asked to phone in their response to a survey question.

Section 1–4

For Exercises 45–48, identify each study as being either observational or experimental.

45. Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After 6 months, the numbers of respiratory tract infections each group had were compared.

46. A researcher stood at a busy intersection to see if the color of the automobile that a person drives is related to running red lights.

47. A researcher finds that people who are more hostile have higher total cholesterol levels than those who are less hostile.

48. Subjects are randomly assigned to four groups. Each group is placed on one of four special diets—a low-fat diet, a high-fish diet, a combination of low-fat diet and high-fish diet, and a regular diet. After 6 months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure.

For Exercises 49–52, identify the independent and dependent variables for each study.

49. Various types of coffees are selected from local coffee shops, and the number of milligrams of caffeine per ounce is determined.

50. People who walk at least 3 miles a day are randomly selected, and their blood triglyceride levels are measured in order to determine if the number of miles that they walk has any influence on these levels.

51. In an article in the *British Journal of Nutrition,* two types of mice were randomly selected. One group received a thyme supplement for a specific time, while another group was used as a control group and received no supplements. The brains of the mice were then analyzed, and it was found that the brains of the group of mice that received the thyme supplements had antioxidant levels similar to those of younger mice. It was concluded that the thyme supplement increased the antioxidants in the brains of the mice.

52. It was found that laughing can reduce your aches and pains as much as taking a pain pill.

For Exercises 53–58, explain why the claims of these studies might be suspect.

53. Based on a recent telephone survey, 72% of those contacted shop online.

54. In Greenville County there are 8324 deer.

55. Nursing school graduates from Fairview University earn on average $33,456.

56. Only 5% of men surveyed said that they liked "chick flicks."

57. A recent study shows that high school dropouts spend less time on the Internet than those who graduated; therefore, the Internet raises your IQ.

58. Most shark attacks occur in ocean water that is 3 feet deep; therefore, it is safer to swim in deep water.

STATISTICS TODAY

Is Higher Education "Going Digital"? —Revisited

Researchers at the Pew Research Center used a telephone survey of 2142 graduates and an online survey of 1055 college and university presidents of two-year and four-year public and private colleges to ascertain their findings.

They found out that approximately 89% of public colleges and universities offer online classes while 60% of four-year private colleges offer them. About 23% of the graduates said that they have taken an online course. The college presidents predict that in 10 years, most of their students will have taken an online course.