



The Nature of Probability and Statistics

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.

Outline

- 1-1 Introduction
- 1-2 Descriptive and Inferential Statistics
- 1-3 Variables and Types of Data
- 1-4 Data Collection and Sampling Techniques
- 1-5 Observational and Experimental Studies
- 1-6 Uses and Misuses of Statistics
- 1-7 Computers and Calculators
- 1-8 Summary



Statistics Today

Are We Improving Our Diet?

It has been determined that diets rich in fruits and vegetables are associated with a lower risk of chronic diseases such as cancer. Nutritionists recommend that Americans consume five or more servings of fruits and vegetables each day. Several researchers from the Division of Nutrition, the National Center for Chronic Disease Control and Prevention, the National Cancer Institute, and the National Institutes of Health decided to use statistical procedures to see how much progress is being made toward this goal.

The procedures they used and the results of the study will be explained in this chapter. See Statistics Today—Revisited at the end of this chapter.

1-1

Introduction

Most people become familiar with probability and statistics through radio, television, newspapers, and magazines. For example, the following statements were found in newspapers.

- Nearly one in seven U.S. families are struggling with bills from medical expenses even though they have health insurance. Source: *Psychology Today*, October 2004.
- Eating 10 grams (g) of fiber a day reduces the risk of heart attack by 14%. Source: Archives of Internal Medicine, *Reader's Digest*, May 2004.
- Thirty minutes (of exercise) two or three times each week can raise HDLs 10 to 15%. Source: *Prevention*, July 2004.
- The average credit card debt per household in 2003 was \$9205. Source: www.cardweb.com.
- About 15% of men in the United States are left-handed and 9% of women are left-handed. Source: Scripps Survey Research Center.
- The median age of couples who watch Jay Leno is 48.1 years. Source: Nielsen Media Research 2003–2004.

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.

Interesting Fact

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

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.

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.

Students study statistics for several reasons:

1. Students, like professional people, must be able to read and understand the various statistical studies performed in their fields. To have this understanding, they must be knowledgeable about the vocabulary, symbols, concepts, and statistical procedures used in these studies.
2. Students and professional people may be called on to conduct research in their fields, since statistical procedures are basic to research. To accomplish this, they must be able to design experiments; collect, organize, analyze, and summarize data; and possibly make reliable predictions or forecasts for future use. They must also be able to communicate the results of the study in their own words.
3. Students and professional people can also use the knowledge gained from studying statistics to become better consumers and citizens. For example, they 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 the goals for students and professionals who study statistics.

It is the purpose of this chapter to introduce the student to the basic concepts of *probability* and *statistics* by answering questions such as the following:

What are the branches of statistics?

What are data?

How are samples selected?

1–2**Descriptive and Inferential Statistics**

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

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

Objective 1

Demonstrate knowledge of statistical terms.

Objective 2

Differentiate between the two branches of statistics.

Historical Note

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

Unusual Stat

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

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**.

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

1. Descriptive statistics
2. Inferential statistics

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 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.

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

The second area of statistics is called *inferential statistics*. Here, the statistician tries to make inferences from *samples* to *populations*. Inferential statistics uses **probability**, i.e., the chance of an event occurring. Many people are familiar with the concepts of probability through various forms of gambling. People who play cards, dice, bingo, and lotteries win or lose according to the laws of probability. Probability theory is also used in the insurance industry and other areas.

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

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

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

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

If the subjects of a sample are properly selected, most of the time they should possess the same or similar characteristics as the subjects in the population. The techniques used to properly select a sample will be explained in Section 1–4.

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

Statisticians also use statistics to determine *relationships* among variables. For example, relationships were the focus of the most noted study in the past few decades,

Speaking of Statistics

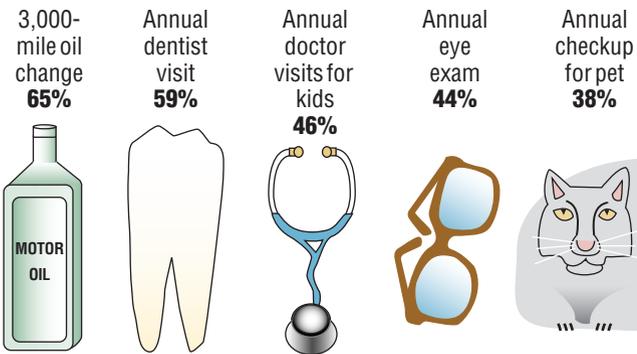
This *USA TODAY* Snapshot shows the percentages of adults who say these activities are extremely important. Based on the information shown in the Snapshot, do you think people take better care of their automobiles than they do of themselves?

USA SNAPSHOTS®

A look at statistics that shape our lives

Medical visits take a back seat

Percentage of adults who say these activities are extremely important:



Source: Roper/ASW for TLC Laser Eye Centers

Source: Copyright 2002, *USA TODAY*. Reprinted with permission.

Historical Note

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

“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.

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.

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

Applying the Concepts 1–2

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.

Based on this information, attendance and grades are related. The more you attend class, the more likely you will receive a higher grade. If you improve your attendance, your grades

Unusual Stat

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

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 32 for the answers.

1–3

Variables and Types of Data

Objective 3

Identify types of data.

As stated in Section 1–2, statisticians gain information about a particular situation by collecting data for random variables. This section will explore in greater detail the nature of variables and types of data.

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

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

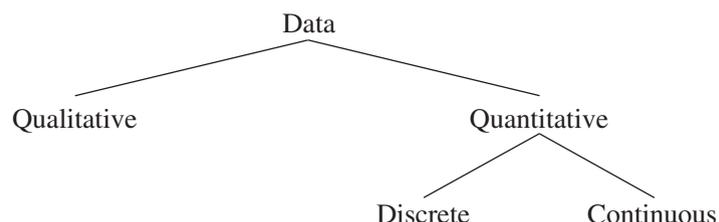
Quantitative variables can be further classified into two groups: discrete and continuous. *Discrete variables* can be assigned values such as 0, 1, 2, 3 and are said to be countable. Examples of discrete variables are the number of children in a family, the number of students in a classroom, and the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

Discrete variables assume values that can be counted.

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

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

The classification of variables can be summarized as follows:



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. *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

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).

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

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.

Unusual Stat

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

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°F 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.

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.

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

Applying the Concepts 1–3

Safe Travel

Read the following information about the transportation industry and answer the questions.

Transportation Safety

The chart shows the number of job-related injuries for each of the transportation industries for 1998.

Industry	Number of injuries
Railroad	4520
Intercity bus	5100
Subway	6850
Trucking	7144
Airline	9950

1. What are 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 is shown as the safest transportation industry. Does that mean railroads have fewer accidents than the other industries? Explain.
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 32 for the answers.

1–4

Objective 5

Identify the four basic sampling techniques.

Data Collection and 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.

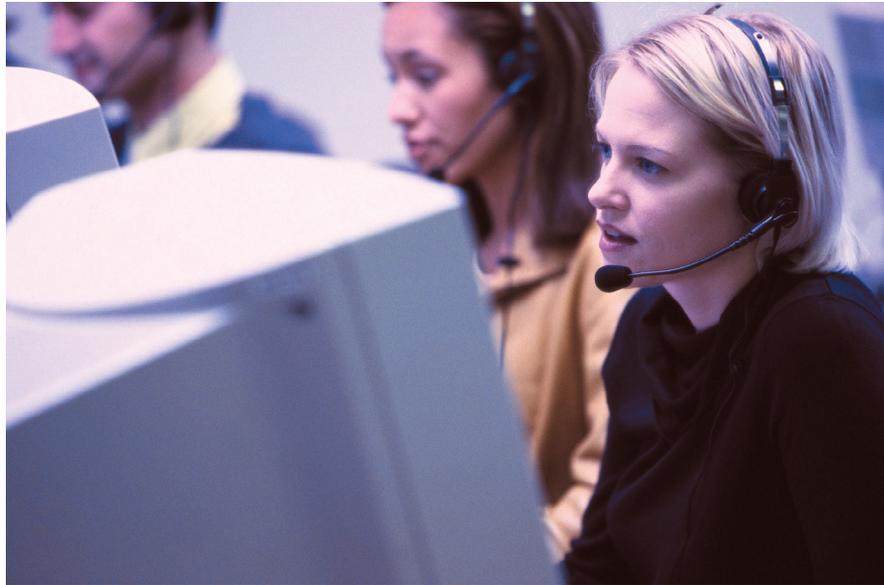
Data can be collected in a variety of ways. One of the most common methods is through the use of surveys. Surveys can be done by using a variety of methods. Three of the most common methods are the telephone survey, the mailed questionnaire, and the personal interview.

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 the 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

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.



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 1–2, 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. 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., 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.

Random Sampling

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 1 to 85.

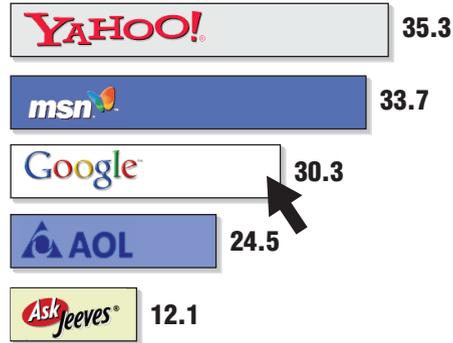
Speaking of Statistics

This study of Internet search sites was conducted by Nielsen/NetRatings. What type of sample do you think they used?

USA SNAPSHOTS®

Top Internet search sites

Users¹ in millions for March:



Source: Nielsen/Net Ratings

¹ – Each visitor is counted only once.

Source: Copyright 2002, USA TODAY. Reprinted with permission.

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

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 C.

Systematic Sampling

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 population and a sample of 50 subjects were 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, one 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 tenth item from an assembly line to test for defects.

Stratified Sampling

Researchers obtain **stratified samples** by dividing the population into groups (called strata) according to some characteristic that is important to the study, then sampling from each group. 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 the first-year students differ from those of the second-year students. The president will select students from each group to use in the sample.

Cluster Sampling

Researchers also use **cluster samples**. Here the population is divided into groups called clusters by some means such as geographic area or schools in a large school district, etc. 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.

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

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.

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 groups (strata), and subjects within groups are randomly selected.
Cluster	Subjects are selected by using an intact group that is representative of the population.

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 65 would.

Other Sampling Methods

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 that 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.

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

Applying the Concepts 1–4

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, 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 page 32 for the answers.

1–5

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*.

Objective 6

Explain the difference between an observational and an experimental study.

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.

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 due to 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

Interesting Fact

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

amount of medicine each day, the researcher can determine the precise dosages and, if necessary, vary the dosage for the groups.

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 is called *confounding of variables*.

A confounding variable is one that influences the dependent or outcome variable but cannot be 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.

Observational studies also 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.

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.

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?

Unusual Stat

Of people in the United States, 66% read the Sunday paper.

The answer here is to take a closer look at these studies. Actually, it is not a choice between butter or 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 to use liquid margarine instead of 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.

Applying the Concepts 1–5

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 been 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 32 for the answers.

1–6

Uses and Misuses of Statistics

Objective 7

Explain how statistics can be used and misused.

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.

There are two sayings that have been around for a long time that 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, one 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. One 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. 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 which lends the most evidence to support their position. This fact is explained on page 105.

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 best 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 false conclusions. The misuse of graphs is also explained in Chapter 2, on pages 70–73.

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.

To restate the premise of this section, 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!

1-7

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-83 Plus or TI-84 Plus graphing calculator accomplishes the same thing.

There are many statistical packages available; this book uses MINITAB and Microsoft Excel. Instructions for using MINITAB, the TI-83 Plus or TI-84 Plus graphing calculator, and Excel have been placed at the end of each relevant section, in subsections entitled Technology Step by Step.

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

MINITAB Step by Step

General Information

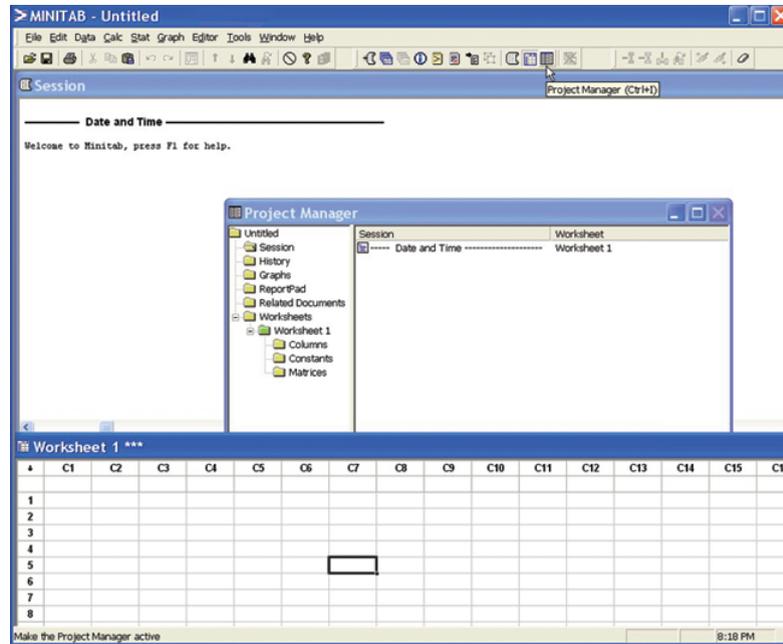
MINITAB statistical software provides a wide range of statistical analysis and graphing capabilities.

Take Note

In this text you will see captured screen images from computers running MINITAB Release 14. If you are using an earlier release of MINITAB, the screens you see on your computer may bear slight visual differences from the screens pictured in this text. **But don't be alarmed!** All the Step by Step operations described in this text, including the commands, the menu options, and the functionality, will work just fine on your computer.

Start the Program

1. Click the Windows XP Start Menu, then All Programs.
2. Click the MINITAB 14 folder and then click  MINITAB 14, 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 in MINITAB

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 one 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.

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. For a floppy disk this would be **A:**.
 - b) **File Name:** Type in the name of the file, such as **MyData**.
 - c) **Save as Type:** 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.MTW**.

Open the Databank File

The raw data are shown in Appendix D. 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

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

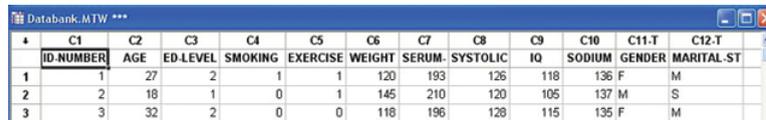


Databank, are available on the accompanying CD-ROM or at the Online Learning Center (www.mhhe.com/bluman). Here is how to get the data from a file into a worksheet.

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 [*.*mtw].
 - c) Double-click the file name in the list box Databank.mtw. 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.



	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.

TI-83 Plus or TI-84 Plus Step by Step

The TI-83 Plus or 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 (if necessary, clear the old list):

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-83 Plus or 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

L_1	L_2	L_3	1
213	-----	-----	
208			
203			
215			
222			

L1(6)=			

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 cursor to incorrect value in list and type in the correct value. Then press **ENTER**.

To delete a data value in a list:

Move cursor to 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 [L₁]**. Then press **ENTER**.

(*Note:* To clear several lists, follow step 1, but enter each list to be cleared, separating them with commas.)

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 [L₁] ENTER**.

The calculator will display Done.

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

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

Example TI1-2

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

Output

L1	L2	L3	1
203	-----	-----	
208			
213			
215			
222			

L1(6)=			

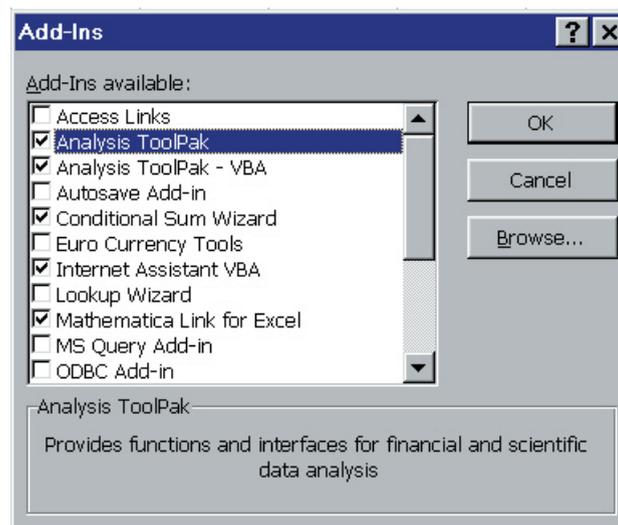
Excel Step by Step

General Information

Microsoft Excel 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 the f_x icon. For most of the problems in this textbook, however, it is easier to use the packaged tests in the **Data Analysis Add-in**.

To activate the **Data Analysis Add-in**:

1. Select **Tools** on the Worksheet menu bar.
2. Select **Add-Ins** from the **Tools** menu.
3. Check the box for **Analysis ToolPak**.



Excel's Analysis
ToolPak Add-In

MegaStat

Later in this text you will encounter a few Excel Technology Step by Step operations that would not be feasible without another Excel add-in: MegaStat. MegaStat is provided on the CD that came with your textbook, as well as on your text's Online Learning Center at www.mhhe.com/bluman.

1. Locate the MegaStat installation program by double-clicking "My Computer" on your desktop and looking in the proper directory. (If you downloaded MegaStat from the OLC, it will be in the location to which you saved it. If you are accessing it from your CD, simply double-click on your CD drive.)
2. Follow the instructions to install MegaStat on your computer.
3. Activate the MegaStat add-in in the same way you activated the Data Analysis add-in: select **Tools>Add-ins**, and check the box for MegaStat.

Entering Data

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

You can also enter data into Excel by opening an Excel workbook file. You can access and download data sets used in this book by using the accompanying CD-ROM or the Online Learning Center (www.mhhe.com/bluman). Excel data are in files called Workbooks. Follow the steps in Example XL1-1 to open a workbook file.

Example XL1-1

1. Click **File** on the menu bar, then click Open . . . (i.e., select **File>Open . . .**). The Open dialog box will be displayed.
2. In the Look in box, click the folder where the workbook file is located.
3. In the folder list, double-click folders until you open the folder containing the file you want.
4. Double-click the file name in the list box. The data will be copied into Excel.

1-8

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*.

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.

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.

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.

Unusual Stat

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

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.

Finally, the applications of statistics are many and varied. People encounter them in everyday life, such as in reading newspapers or magazines, listening to the radio, 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.

Today, computers and calculators are used extensively in statistics to facilitate the computations.



Source: © 1993 King Features Syndicate, Inc. World Rights reserved. Reprinted with special permission of King Features Syndicate.

Important Terms

cluster sample	12	experimental study	14	observational study	13	random variable	3
confounding variable	15	explanatory variable	14	ordinal level of measurement	8	ratio level of measurement	8
continuous variables	6	Hawthorne effect	15	outcome variable	14	sample	4
control group	14	hypothesis testing	4	population	4	statistics	3
convenience sample	13	independent variable	14	probability	4	stratified sample	12
data	3	inferential statistics	5	qualitative variables	6	systematic sample	12
data set	4	interval level of measurement	8	quantitative variables	6	treatment group	14
data value or datum	4	measurement scales	7	quasi-experimental study	14	variable	3
dependent variable	14	nominal level of measurement	7	random sample	10		
descriptive statistics	4						
discrete variables	6						

Review Exercises

Note: All odd-numbered problems and even-numbered problems marked with “ans” are included in the answer section at the end of this book.

- Name and define the two areas of statistics.
- What is probability? Name two areas where probability is used.
- Suggest some ways statistics can be used in everyday life.
- Explain the differences between a sample and a population.
- Why are samples used in statistics?
- (ans)** In each of these statements, tell whether descriptive or inferential statistics have been used.
 - In the year 2010, 148 million Americans will be enrolled in an HMO (Source: *USA TODAY*).
 - Nine out of ten on-the-job fatalities are men (Source: *USA TODAY Weekend*).
 - Expenditures for the cable industry were \$5.66 billion in 1996 (Source: *USA TODAY*).
 - The median household income for people aged 25–34 is \$35,888 (Source: *USA TODAY*).
 - Allergy therapy makes bees go away (Source: *Prevention*).
 - Drinking decaffeinated coffee can raise cholesterol levels by 7% (Source: American Heart Association).
 - The national average annual medicine expenditure per person is \$1052 (Source: *The Greensburg Tribune Review*).
 - Experts say that mortgage rates may soon hit bottom (Source: *USA TODAY*).
- Classify each as nominal-level, ordinal-level, interval-level, or ratio-level measurement.
 - Pages in the city of Cleveland telephone book.
 - Rankings of tennis players.
 - Weights of air conditioners.
 - Temperatures inside 10 refrigerators.
 - Salaries of the top five CEOs in the United States.
 - Ratings of eight local plays (poor, fair, good, excellent).
 - Times required for mechanics to do a tune-up.
 - Ages of students in a classroom.
 - Marital status of patients in a physician’s office.
 - Horsepower of tractor engines.
- (ans)** Classify each variable as qualitative or quantitative.
 - Number of bicycles sold in 1 year by a large sporting goods store.
 - Colors of baseball caps in a store.
 - Times it takes to cut a lawn.
 - Capacity in cubic feet of six truck beds.
 - Classification of children in a day-care center (infant, toddler, preschool).
 - Weights of fish caught in Lake George.
 - Marital status of faculty members in a large university.
- Classify each variable as discrete or continuous.
 - Number of doughnuts sold each day by Doughnut Heaven.
 - Water temperatures of six swimming pools in Pittsburgh on a given day.
 - Weights of cats in a pet shelter.
 - Lifetime (in hours) of 12 flashlight batteries.
 - Number of cheeseburgers sold each day by a hamburger stand on a college campus.
 - Number of DVDs rented each day by a video store.
 - Capacity (in gallons) of six reservoirs in Jefferson County.
- Give the boundaries of each value.
 - 42.8 miles.
 - 1.6 milliliters.
 - 5.36 ounces.
 - 18 tons.
 - 93.8 ounces.
 - 40 inches.
- Name and define the four basic sampling methods.
- (ans)** Classify each sample as random, systematic, stratified, or cluster.
 - 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.
 - Every seventh customer entering a shopping mall is asked to select her or his favorite store.
 - Nursing supervisors are selected using random numbers in order to determine annual salaries.
 - Every 100th hamburger manufactured is checked to determine its fat content.
 - 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.
13. Give three examples each of nominal, ordinal, interval, and ratio data.
 14. For each of these statements, define a population and state how a sample might be obtained.
 - a. The average cost of an airline meal is \$4.55 (Source: *Everything Has Its Price*, Richard E. Donley, Simon and Schuster).
 - b. More than 1 in 4 United States children have cholesterol levels of 180 milligrams or higher (Source: The American Health Foundation).
 - c. Every 10 minutes, 2 people die in car crashes and 170 are injured (Source: National Safety Council estimates).
 - d. When older people with mild to moderate hypertension were given mineral salt for 6 months, the average blood pressure reading dropped by 8 points systolic and 3 points diastolic (Source: *Prevention*).
 - e. The average amount spent per gift for Mom on Mother's Day is \$25.95 (Source: The Gallup Organization).
 15. 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.
 16. Information from research studies is sometimes taken out of context. Explain why the claims of these studies might be suspect.
 - a. The average salary of the graduates of the class of 1980 is \$32,500.
 - b. It is estimated that in Podunk there are 27,256 cats.
 - c. Only 3% of the men surveyed read *Cosmopolitan* magazine.
 - d. Based on a recent mail survey, 85% of the respondents favored gun control.
 - e. A recent study showed that high school dropouts drink more coffee than students who graduated; therefore, coffee dulls the brain.
 - f. Since most automobile accidents occur within 15 miles of a person's residence, it is safer to make long trips.
 17. Identify each study as being either observational or experimental.
 - a. 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.
 - b. 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.
 - c. A researcher finds that people who are more hostile have higher total cholesterol levels than those who are less hostile.
 - d. 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.
 18. Identify the independent variable(s) and the dependent variable for each of the studies in Exercise 17.
 19. For each of the studies in Exercise 17, suggest possible confounding variables.
 20. 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?
 21. 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!"
 22. 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.
 23. 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?

24. In an ad for moisturizing lotion, the following claim is made: “. . . it’s the #1 dermatologist-recommended brand.” What is misleading about this claim?
25. An ad for an exercise product stated: “Using this product will burn 74% more calories.” What is misleading about this statement?
26. “Vitamin E is a proven antioxidant and may help in fighting cancer and heart disease.” Is there anything ambiguous about this claim? Explain.
27. “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?
28. “. . . 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?
29. 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

30. Find an article that describes a statistical study, and identify the study as observational or experimental.
31. For the article that you used in Exercise 30, identify the independent variable(s) and dependent variable for the study.
32. For the article that you selected in Exercise 30, suggest some confounding variables that may have an effect on the results of the study.

Statistics Today

Are We Improving Our Diet?—Revisited

Researchers selected a *sample* of 23,699 adults in the United States, using phone numbers selected at *random*, and conducted a *telephone survey*. All respondents were asked six questions:

1. How often do you drink juices such as orange, grapefruit, or tomato?
2. Not counting juice, how often do you eat fruit?
3. How often do you eat green salad?
4. How often do you eat potatoes (not including french fries, fried potatoes, or potato chips)?
5. How often do you eat carrots?
6. Not counting carrots, potatoes, or salad, how many servings of vegetables do you usually eat?

Researchers found that men consumed fewer servings of fruits and vegetables per day (3.3) than women (3.7). Only 20% of the population consumed the recommended 5 or more daily servings. In addition, they found that youths and less-educated people consumed an even lower amount than the average.

Based on this study, they recommend that greater educational efforts are needed to improve fruit and vegetable consumption by Americans and to provide environmental and institutional support to encourage increased consumption.

Source: Mary K. Serdula, M.D., et al., “Fruit and Vegetable Intake Among Adults in 16 States: Results of a Brief Telephone Survey,” *American Journal of Public Health* 85, no. 2. Copyright by the American Public Health Association.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

- Probability is used as a basis for inferential statistics.
- The height of President Lincoln is an example of a variable.
- The highest level of measurement is the interval level.
- When the population of college professors is divided into groups according to their rank (instructor, assistant professor, etc.) and then several are selected from each group to make up a sample, the sample is called a cluster sample.
- The variable *age* is an example of a qualitative variable.
- The weight of pumpkins is considered to be a continuous variable.
- The boundary of a value such as 6 inches would be 5.9–6.1 inches.

Select the best answer.

- The number of absences per year that a worker has is an example of what type of data?
 - Nominal
 - Qualitative
 - Discrete
 - Continuous
- What are the boundaries of 25.6 ounces?
 - 25–26 ounces
 - 25.55–25.65 ounces
 - 25.5–25.7 ounces
 - 20–39 ounces
- A researcher divided subjects into two groups according to gender and then selected members from each group for her sample. What sampling method was the researcher using?
 - Cluster
 - Random
 - Systematic
 - Stratified
- Data that can be classified according to color are measured on what scale?
 - Nominal

- Ratio
- Ordinal
- Interval

- A study that involves no researcher intervention is called
 - An experimental study.
 - A noninvolvement study.
 - An observational study.
 - A quasi-experimental study.
- A variable that interferes with other variables in the study is called
 - A confounding variable.
 - An explanatory variable.
 - An outcome variable.
 - An interfering variable.

Use the best answer to complete these statements.

- Two major branches of statistics are _____ and _____.
- Two uses of probability are _____ and _____.
- The group of all subjects under study is called a(n) _____.
- A group of subjects selected from the group of all subjects under study is called a(n) _____.
- Three reasons why samples are used in statistics are
 - _____
 - _____
 - _____.
- The four basic sampling methods are
 - _____
 - _____
 - _____
 - _____.
- A study that uses intact groups when it is not possible to randomly assign participants to the groups is called a(n) _____ study.
- In a research study, participants should be assigned to groups using _____ methods, if possible.
- For each statement, decide whether descriptive or inferential statistics is used.
 - The average life expectancy in New Zealand is 78.49 years. Source: *World Factbook 2004*.
 - A diet high in fruits and vegetables will lower blood pressure. Source: Institute of Medicine.
 - The total amount of estimated losses from hurricane Hugo was \$4.2 billion. Source: Insurance Service Office.

- d. Researchers stated that the shape of a person's ears is related to the person's aggression. Source: *American Journal of Human Biology*.
- e. In 2013, the number of high school graduates will be 3.2 million students. Source: National Center for Education.
23. Classify each as nominal-level, ordinal-level, interval-level, or ratio-level measurement.
- Rating of movies as G, PG, and R.
 - Number of candy bars sold on a fund drive.
 - Classification of automobiles as subcompact, compact, standard, and luxury.
 - Temperatures of hair dryers.
 - Weights of suitcases on a commercial airline.
24. Classify each variable as discrete or continuous.
- Ages of people working in a large factory.
 - Number of cups of coffee served at a restaurant.
 - The amount of a drug injected into a guinea pig.
 - The time it takes a student to drive to school.
 - The number of gallons of milk sold each day at a grocery store.
25. Give the boundaries of each.
- 48 seconds
 - 0.56 centimeter
 - 9.1 quarts
 - 13.7 pounds
 - 7 feet

Critical Thinking Challenges

- A study of the world's busiest airports was conducted by *Airports Council International*. Describe three variables that one could use to determine which airports are the busiest. What *units* would one use to measure these variables? Are these variables categorical, discrete, or continuous?
- The results of a study published in *Archives of General Psychiatry* stated that male children born to women who smoke during pregnancy run a risk of violent and criminal behavior that lasts into adulthood. The results of this study were challenged by some people in the media. Give several reasons why the results of this study would be challenged.
- The results of a study published in *Neurological Research* stated that second-graders who took piano lessons and played a computer math game more readily grasped math problems in fractions and proportions than a similar group who took an English class and played the same math game. What type of inferential study was this? Give several reasons why the piano lessons could improve a student's math ability.
- A study of 2958 collegiate soccer players showed that in 46 anterior cruciate ligament (ACL) tears, 36 were in women. Calculate the percentages of tears for each gender.
 - Can it be concluded that female athletes tear their knees more often than male athletes?
 - Comment on how this study's conclusion might have been reached.
- Read the article entitled "Anger Can Cause Snap Judgments" and answer the following questions.
 - Is the study experimental or observational?
 - What is the independent variable?
 - What is the dependent variable?
 - Do you think the sample sizes are large enough to merit the conclusion?
 - Based on the results of the study, what changes would you recommend to persons to help them reduce their anger?
- Read the article entitled "Hostile Children Fight Unemployment" and answer the following questions.
 - Is the study experimental or observational?
 - What is the independent variable?
 - What is the dependent variable?
 - Suggest some confounding variables that may have influenced the results of the study.
 - Identify the three groups of subjects used in the study.

ANGER CAN CAUSE SNAP JUDGMENTS

Anger can make a normally unbiased person act with prejudice, according to a forthcoming study in the journal *Psychological Science*.

Assistant psychology professors David DeSteno at Northeastern University in Boston and Nilanjana Dasgupta at the University of Massachusetts, Amherst, randomly divided 81 study participants into two groups and assigned them a writing task designed to induce angry, sad or neutral feelings. In a subsequent test to uncover nonconscious associations, angry subjects were quicker to

connect negatively charged words—like war, death and vomit—with members of the opposite group—even though the groupings were completely arbitrary.

“These automatic responses guide our behavior when we’re not paying attention,” says DeSteno, and they can lead to discriminatory acts when there is pressure to make a quick decision. “If you’re aware that your emotions might be coloring these gut reactions,” he says, “you should take time to consider that possibility and adjust your actions accordingly.”

—Eric Strand

Source: Reprinted with permission from *Psychology Today*, Copyright © (2004) Sussex Publishers, Inc.

UNEMPLOYMENT

Hostile Children Fight Unemployment

Aggressive children may be destined for later long-term unemployment. In a study that began in 1968, researchers at the University of Jyväskylä in Finland examined about 300 participants at ages 8, 14, 27, and 36. They looked for aggressive behaviors like hurting other children, kicking objects when angry, or attacking others without reason.

Their results, published recently in the *International Journal of Behavioral Development*, suggest that children with low self-control of emotion—especially aggression—were significantly more prone to long-term unemployment. Children with behavioral inhibitions—such as passive and anxious behaviors—were also indirectly linked to unemployment

as they lacked the preliminary initiative needed for school success. And while unemployment rates were high in Finland during the last data collection, jobless participants who were aggressive as children were less likely to have a job two years later than their nonaggressive counterparts.

Ongoing unemployment can have serious psychological consequences, including depression, anxiety and stress. But lead researcher Lea Pulkkinen, Ph.D., a Jyväskylä psychology professor, does have encouraging news for parents: Aggressive children with good social skills and child-centered parents were significantly less likely to be unemployed for more than two years as adults.

—Tanya Zimbardo

Source: Reprinted with permission from *Psychology Today*, Copyright © (2001) Sussex Publishers, Inc.

Answers to Applying the Concepts

Section 1–2 Attendance and Grades

1. The variables are grades and attendance.
2. The data consist of specific grades and attendance numbers.
3. These are descriptive statistics.
4. The population under study is students at Manatee Community College (MCC).
5. While not specified, we probably have data from a sample of MCC students.
6. Based on the data, it appears that in general, the better your attendance the higher your grade.

Section 1–3 Safe Travel

1. The variables are industry and number of job-related injuries.
2. The type of industry is a qualitative variable, while the number of job-related injuries is quantitative.
3. The number of job-related injuries is discrete.
4. Type of industry is nominal, and the number of job-related injuries is ratio.
5. The railroads do show fewer job-related injuries; however, there may be other things to consider. For example, railroads employ fewer people than the other transportation industries in the study.
6. A person's choice of transportation might also be affected by convenience issues, cost, service, etc.
7. Answers will vary. One possible answer is that the railroads have the fewest job-related injuries, while the airline industry has the most job-related industries (more than twice those of the railroad industry). The numbers of job-related industries in the subway and trucking industries are fairly comparable.

Section 1–4 American Culture and Drug Abuse

Answers will vary, so this is one possible answer.

1. I used a telephone survey. The advantage to my survey method is that this was a relatively inexpensive survey method (although more expensive than using the mail) that could get a fairly sizable response. The disadvantage to my survey method is that I have not included anyone without a telephone. (*Note:* My survey used a random dialing method to include unlisted numbers and cell phone exchanges.)
2. A mail survey also would have been fairly inexpensive, but my response rate may have been much lower than

what I got with my telephone survey. Interviewing would have allowed me to use follow-up questions and to clarify any questions of the respondents at the time of the interview. However, interviewing is very labor- and cost-intensive.

3. I used ordinal data on a scale of 1 to 5. The scores were 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.
4. The random method that I used was a random dialing method.
5. To include people from each state, I used a stratified random sample, collecting data randomly from each of the area codes and telephone exchanges available.
6. This method allowed me to make sure that I had representation from each area of the United States.
7. Convenience samples may not be representative of the population, and a convenience sample of adolescents would probably differ greatly from the general population with regard to the influence of American culture on illegal drug use.

Section 1–5 Just a Pinch Between Your Cheek and Gum

1. This was an experiment, since the researchers imposed a treatment on each of the two groups involved in the study.
2. The independent variable is whether the participant chewed tobacco or not. The dependent variables are the students' blood pressures and heart rates.
3. The treatment group is the tobacco group—the other group was used as a control.
4. A student's blood pressure might not be affected by knowing that he or she was part of a study. However, if the student's blood pressure were affected by this knowledge, all the students (in both groups) would be affected similarly. This might be an example of the placebo effect.
5. Answers will vary. One possible answer is that confounding variables might include the way that the students chewed the tobacco, whether or not the students smoked (although this would hopefully have been evened out with the randomization), and that all the participants were university students.
6. Answers will vary. One possible answer is that the study design was fine, but that it cannot be generalized beyond the population of university students (or people around that age).