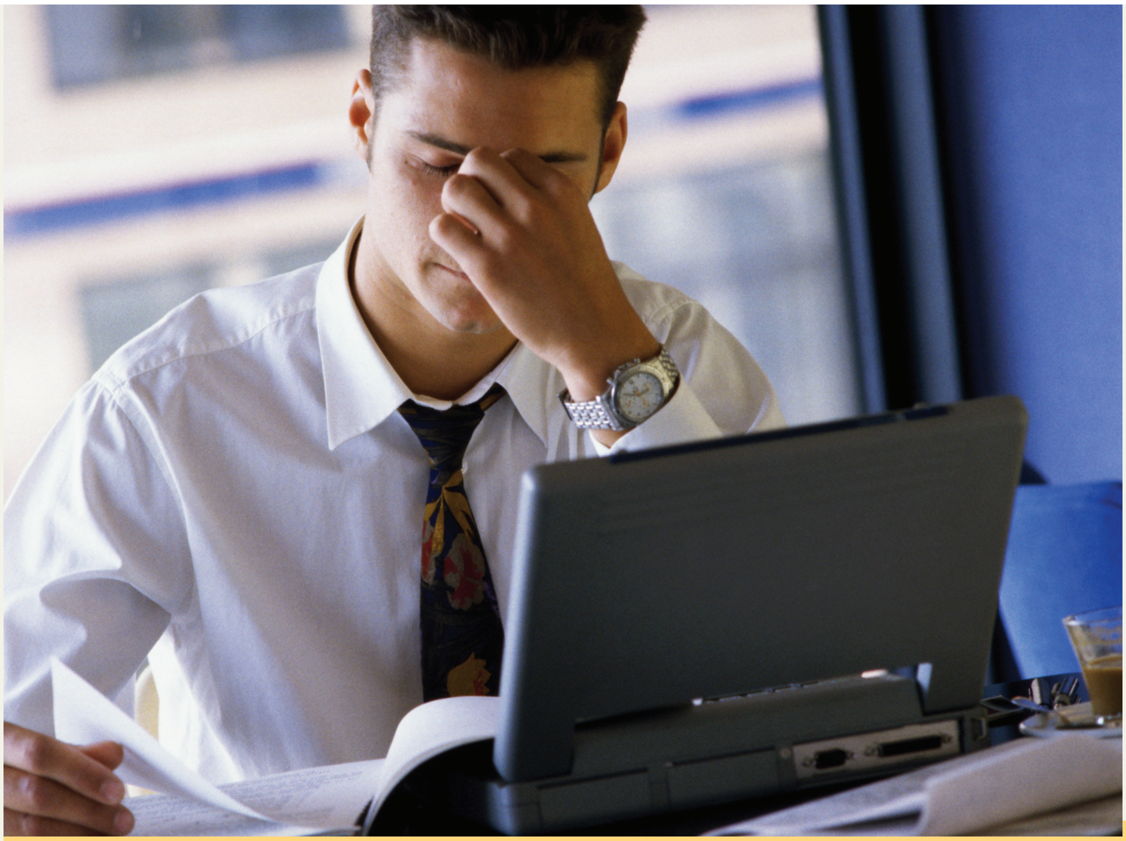


Chapter

1

Introduction

Do you feel pressured by your boss or co-workers to come to work when you are sick with flu? Or are they more considerate and supportive and you do not feel this pressure? In a sample survey, 38% of the respondents said they feel pressured to come to work when they have flu. However, a larger majority, 61%, said that they do not feel pressured. The remaining 1% were not sure. (See Case Study 1–2.)

The study of statistics has become more popular than ever over the past four decades or so. The increasing availability of computers and statistical software packages has enlarged the role of statistics as a tool for empirical research. As a result, statistics is used for research in almost all professions, from medicine to sports. Today, college students in almost all disciplines are required to take at least one statistics course. Almost all newspapers and magazines these days contain graphs and stories on statistical studies. After you finish reading this book, it should be much easier to understand these graphs and stories.

Every field of study has its own terminology. Statistics is no exception. This introductory chapter explains the basic terms of statistics. These terms will bridge our understanding of the concepts and techniques presented in subsequent chapters.

1.1 What is Statistics?

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

After winning the World Series in 1918, the Boston Red Sox did not win the World Series again until 2004. The following table compares a few facts for the years 1918 and 2004. The numbers given in this table can be referred to as *statistics.*

Item	1918	2004
Price of a World Series Ticket	$3.30	$140
U.S. Population (in Millions)	106	294
Price of a Quart of Milk	14 Cents	$1.09
Price of a Loaf of Bread	10 Cents	$2.19

Source: USA TODAY, October 28, 2004.

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

Definition

Statistics *Statistics* is a group of methods used to collect, analyze, present, and interpret data and to make decisions.

Every day we make decisions that may be personal, business related, or of some other kind. Usually these decisions are made under conditions of uncertainty. Many times, the situations or problems we face in the real world have no precise or definite solution. Statistical methods help us make scientific and intelligent decisions in such situations. Decisions made by using statistical methods are called *educated guesses.* Decisions made without using statistical (or scientific) methods are *pure guesses* and, hence, may prove to be unreliable. For example, opening a large store in an area with or without assessing the need for it may affect its success.

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

1.2 Types of Statistics

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

[1]"The Numbers Racket: How Polls and Statistics Lie," *U.S. News & World Report*, July 11, 1988, pp. 44–47.

1–1 CASE STUDY

10 WEEKS IN FRONT OF THE TV

USA TODAY Snapshots

Source: Statistical Abstract of the United States, 2003 By Karl Gelles, USA TODAY

As the accompanying chart shows, Americans spent an average of 1669 hours, which is equivalent to almost 70 days or 10 weeks, watching television in 2004. Note that this was an estimate for 2004 made in 2003. This chart describes the data on television watching by using one number–the average. We will learn about the average in Chapter 3.

Source: *USA TODAY*, March 30, 2004. Copyright © 2004, *USA TODAY*. Chart reproduced with permission.

1.2.1 Descriptive Statistics

Suppose we have information on the test scores of students enrolled in a statistics class. In statistical terminology, the whole set of numbers that represents the scores of students is called a **data set**, the name of each student is called an **element**, and the score of each student is called an **observation**. (These terms are defined in more detail in Section 1.4.)

A data set in its original form is usually very large. Consequently, such a data set is not very helpful in drawing conclusions or making decisions. It is easier to draw conclusions from summary tables and diagrams than from the original version of a data set. So, we reduce data to a manageable size by constructing tables, drawing graphs, or calculating summary measures such as averages. The portion of statistics that helps us do this type of statistical analysis is called **descriptive statistics**.

Definition

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

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

Case Study 1–1 on this page presents an example of descriptive statistics.

1.2.2 Inferential Statistics

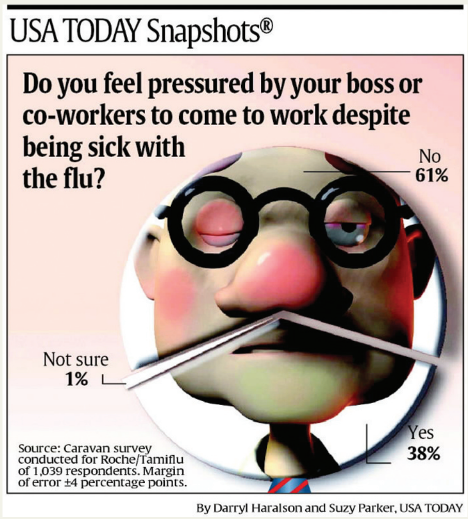
In statistics, the collection of all elements of interest is called a **population**. The selection of a few elements from this population is called a **sample**. (Population and sample are discussed in more detail in Section 1.3.)

A major portion of statistics deals with making decisions, inferences, predictions, and forecasts about populations based on results obtained from samples. For example, we may make some

CASE STUDY

1–2

DO YOU FEEL PRESSURED TO COME TO WORK DESPITE BEING SICK?



The accompanying chart, reproduced from *USA TODAY,* shows that 61% of workers surveyed said that they are not pressured by their boss or co-workers to come to work when they are sick with flu, 38% said they feel pressured, and 1% said that they are not sure. The chart mentions that there is a ±4% margin of error. We will discuss the concept of margin of error in Chapter 8. But just to give a quick and brief explanation, the margin of error means that the percentages given in the chart can change in the plus or minus direction by 4% when applied to the population.

Source: USA TODAY, February 28, 2005. Copyright © 2005, *USA TODAY.* Chart reproduced with permission.

decisions about the political views of all college and university students based on the political views of 1000 students selected from a few colleges and universities. As another example, we may want to find the starting salary of a typical college graduate. To do so, we may select 2000 recent college graduates, find their starting salaries, and make a decision based on this information. The area of statistics that deals with such decision-making procedures is referred to as **inferential statistics**. This branch of statistics is also called *inductive reasoning* or *inductive statistics.*

Definition

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

Case Study 1–2 presents an example of inferential statistics. It shows the results of a survey in which people were asked whether or not they are pressured to come to work despite being sick with flu.

Chapters 8 through 13 and parts of Chapter 7 deal with inferential statistics.

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

EXERCISES

CONCEPTS AND PROCEDURES

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

1.2 Briefly explain the types of statistics.

1.3 Population Versus Sample

We will encounter the terms *population* and *sample* on almost every page of this text.[2] Consequently, understanding the meaning of each of these two terms and the difference between them is crucial.

Suppose a statistician is interested in knowing

1. The percentage of all voters in a city who will vote for a particular candidate in an election
2. The 2005 gross sales of all companies in New York City
3. The prices of all houses in California

In these examples, the statistician is interested in *all* voters, *all* companies, and *all* houses. Each of these groups is called the population for the respective example. In statistics, a population does not necessarily mean a collection of people. It can, in fact, be a collection of people or of any kind of item such as houses, books, television sets, or cars. The population of interest is usually called the **target population**.

Definition

Population or Target Population A *population* consists of all elements—individuals, items, or objects—whose characteristics are being studied. The population that is being studied is also called the *target population.*

Most of the time, decisions are made based on portions of populations. For example, the election polls conducted in the United States to estimate the percentages of voters who favor various candidates in any presidential election are based on only a few hundred or a few thousand voters selected from across the country. In this case, the population consists of all registered voters in the United States. The sample is made up of a few hundred or few thousand voters who are included in an opinion poll. Thus, the collection of a few elements selected from a population is called a **sample**.

Definition

Sample A portion of the population selected for study is referred to as a *sample.*

Figure 1.1 illustrates the selection of a sample from a population.

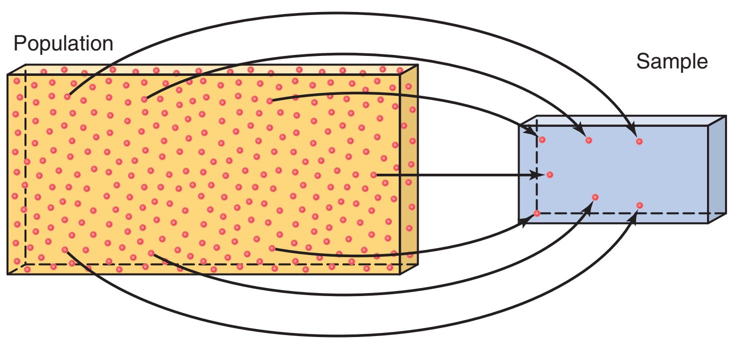


Figure 1.1 Population and sample.

[2]To learn more about sampling and sampling techniques, refer to Appendix A in this text.

The collection of information from the elements of a population or a sample is called a **survey**. A survey that includes every element of the target population is called a **census**. Often the target population is very large. Hence, in practice, a census is rarely taken because it is expensive and time-consuming. In many cases, it is even impossible to identify each element of the target population. Usually, to conduct a survey, we select a sample and collect the required information from the elements included in that sample. We then make decisions based on this sample information. Such a survey conducted on a sample is called a **sample survey**. As an example, if we collect information on the 2005 incomes of all families in Connecticut, it will be referred to as a census. On the other hand, if we collect information on the 2005 incomes of 50 families from Connecticut, it will be called a sample survey.

Definition

Census and Sample Survey A survey that includes every member of the population is called a *census*. The technique of collecting information from a portion of the population is called a *sample survey*.

Case Study 1–3 on page 7 presents an example of a sample survey.

The purpose of conducting a sample survey is to make decisions about the corresponding population. It is important that the results obtained from a sample survey closely match the results that we would obtain by conducting a census. Otherwise, any decision based on a sample survey will not apply to the corresponding population. As an example, to find the average income of families living in New York City by conducting a sample survey, the sample must contain families who belong to different income groups in almost the same proportion as they exist in the population. Such a sample is called a **representative sample**. Inferences derived from a representative sample will be more reliable.

Definition

Representative Sample A sample that represents the characteristics of the population as closely as possible is called a *representative sample.*

A sample may be random or nonrandom. In a **random sample**, each element of the population has a chance of being included in the sample. However, in a nonrandom sample this may not be the case.

Definition

Random Sample A sample drawn in such a way that each element of the population has a chance of being selected is called a *random sample.* If all samples of the same size selected from a population have the same chance of being selected, we call it **simple random sampling**. Such a sample is called a **simple random sample**.

One way to select a random sample is by lottery or draw. For example, if we are to select 5 students from a class of 50, we write each of the 50 names on a separate piece of paper. Then we place all 50 slips in a box and mix them thoroughly. Finally, we randomly draw five slips from the box. The five names drawn give a random sample. On the other hand, if we arrange all 50 names alphabetically and then select the first 5 names on the list, it is a nonrandom sample because the students listed sixth to fiftieth have no chance of being included in the sample.

A sample may be selected with or without replacement. In sampling **with replacement**, each time we select an element from the population, we put it back in the population before we

ELECTRONIC WORLD SWALLOWS UP KIDS' TIME, STUDY FINDS

Children plugged in about 6½ hours a day

The USA's children live in an increasingly heavy stew of media, spending about 6½ hours a day mostly watching TV, using computers and enjoying other electronic activities. And they are spending relatively little time reading or doing homework, a Kaiser Family Foundation survey reported Wednesday.

Kids watch about the same amount of TV—nearly four hours a day—as they did based on a Kaiser survey five years ago, but they're adding newer technology to the mix, such as downloading music and instant-messaging. When multitasking is factored in, children are exposed to 8½ hours of media a day, up about an hour from five years ago.

A record 68% have TVs in their rooms, and an increasing number own DVD and video game players, according to the survey of 2,000 children in grades three through 12.

"We have changed our children's bedrooms into little media arcades," survey co-director Donald Roberts of Stanford University says. "When I was a child, 'Go to your room' was punishment. Now it's 'Go to your room and have a ball.' "

Children with TVs in their rooms watch about 90 minutes more a day and do less reading and homework than those without their own TVs. About half say their families have no TV rules; if there are limits, they're usually not enforced.

"It's alarming, because parents . . . should be setting clear rules and monitoring media use," says Bridget Maher of the Family Research Council, a self-described conservative public policy group.

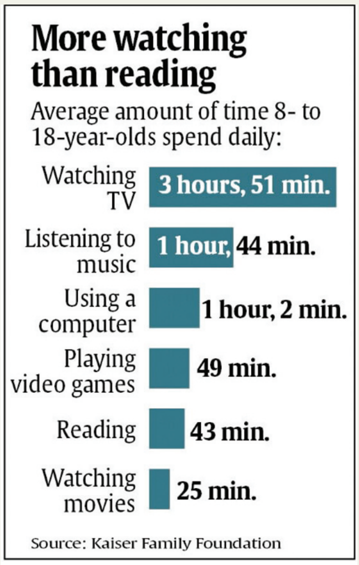
The survey results come amid concern about the soaring rate of childhood obesity. The more kids watch TV, the more likely they are to be heavy, other studies have shown.

Glorified violence on TV is another concern, Roberts says. And new research suggests that violent video games might be even more likely than TV to spur aggression.

But even more serious could be changes in still-developing brains from the constant multitasking, says psychologist Jane Healy, author of *Endangered Minds.* "When you divide attention like this, it becomes harder to focus deeply on any one thing. They may develop habits of mind that make it hard to do in-depth thinking."

The survey underestimates multi-tasking because kids typically have four screens open on a computer, adds Sherry Turkle, an MIT expert on how technology affects people.

But media psychologist Stuart Fischoff of California State University-Los Angeles says all this concern is "premature hysteria." He says he has seen no changes in students' critical thinking during 38 years as a professor, "and TV and the Internet have been around long enough that it would show up by now."



By Joni Alexander, USA TODAY

Source: Marilyn Elias, *USA TODAY,* March 10, 2005. Reproduced with Permission.

select the next element. Thus, in sampling with replacement, the population contains the same number of items each time a selection is made. As a result, we may select the same item more than once in such a sample. Consider a box that contains 25 balls of different colors. Suppose we draw a ball, record its color, and put it back in the box before drawing the next ball. Every time we draw a ball from this box, the box contains 25 balls. This is an example of sampling with replacement.

Sampling **without replacement** occurs when the selected element is not replaced in the population. In this case, each time we select an item, the size of the population is reduced by one element. Thus, we cannot select the same item more than once in this type of sampling. Most of the time, samples taken in statistics are without replacement. Consider an opinion poll based on a certain number of voters selected from the population of all eligible voters. In this case, the same voter is not selected more than once. Therefore, this is an example of sampling without replacement.

EXERCISES

CONCEPTS AND PROCEDURES

1.3 Briefly explain the terms *population*, *sample*, *representative sample*, *random sample*, *sampling with replacement*, and *sampling without replacement.*

1.4 Give one example each of sampling with and sampling without replacement.

1.5 Briefly explain the difference between a census and a sample survey. Why is conducting a sample survey preferable to conducting a census?

APPLICATIONS

1.6 Explain whether each of the following constitutes a population or a sample.
- **a.** Pounds of bass caught by all participants in a bass fishing derby
- **b.** Credit card debts of 100 families selected from a city
- **c.** Number of home runs hit by all Major League baseball players in the 2005 season
- **d.** Number of parole violations by all 2147 parolees in a city
- **e.** Amount spent on prescription drugs by 200 senior citizens in a large city

1.7 Explain whether each of the following constitutes a population or a sample.
- **a.** Number of personal fouls committed by all NBA players during the 2005–2006 season
- **b.** Yield of potatoes per acre for 10 pieces of land
- **c.** Weekly salaries of all employees of a company
- **d.** Cattle owned by 100 farmers in Iowa
- **e.** Number of computers sold during the past week at all computer stores in Los Angeles

1.4 Basic Terms

It is very important to understand the meaning of some basic terms that will be used frequently in this text. This section explains the meaning of an element (or member), a variable, an observation, and a data set. An element and a data set were briefly defined in Section 1.2. This section defines these terms formally and illustrates them with the help of an example.

Table 1.1 gives information on the 2004 profits (in millions of U.S. dollars) of seven U.S. companies. We can call this group of companies a sample of seven companies. Each company listed in this table is called an **element** or a **member** of the sample. Table 1.1 contains information on seven elements. Note that elements are also called *observational units.*

Definition

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

Table 1.1 **2004 Profits of Seven U.S. Companies**

Company	2004 Profits (millions of dollars) ⟵ Variable
Wal-Mart Stores	10,267
Exxon	25,330
An element or a member ⟶ General Electric	16,593 ⟵ An observation or measurement
Citigroup	17,046
Home Depot	5001
Pfizer	11,361
Target	3198

The *2004 profits* in our example is called a **variable**. The *2004 profits* is a characteristic of companies that we are investigating or studying.

Definition

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

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

In general, a variable assumes different values for different elements, as does the 2004 profits of the seven companies in Table 1.1. For some elements in a data set, however, the value of the variable may be the same. For example, if we collect information on incomes of households, these households are expected to have different incomes, although some of them may have the same income.

A variable is often denoted by x, y, or z. For instance, in Table 1.1, the 2004 profits of companies may be denoted by any one of these letters. Starting with Section 1.8, we will begin to use these letters to denote variables.

Each of the values representing the 2004 profits of the seven companies in Table 1.1 is called an **observation** or **measurement**.

Definition

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

From Table 1.1, the 2004 profits of General Electric were $16,593 million. The value $16,593 million is an observation or measurement. Table 1.1 contains seven observations, one for each of the seven companies.

The information given in Table 1.1 on 2004 profits of companies is called the **data** or a **data set**.

Definition

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

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

EXERCISES

CONCEPTS AND PROCEDURES

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

APPLICATIONS

1.9 The following table gives the number of dog bites reported to the police last year in six cities.

City	Number of Bites
Center City	47
Elm Grove	32
Franklin	51
Bay City	44
Oakdale	12
Sand Point	3

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

1.10 The following table lists the crude oil reserves (in billions of barrels) for six countries with the largest reserves as of June 2004.

Country	Oil Reserves
Saudi Arabia	261.7
Iraq	112.0
Kuwait	97.7
Iran	94.4
United Arab Emirates	80.3
Venezuela	64.0

Sources: CIA World Fact Book and *USA TODAY* Research, *USA TODAY,* June 7, 2004.

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

1.11 Refer to the data set in Exercise 1.9.

a. What is the variable for this data set?
b. How many observations are in this data set?
c. How many elements does this data set contain?

1.12 Refer to the data set in Exercise 1.10.

a. What is the variable for this data set?
b. How many observations are in this data set?
c. How many elements does this data set contain?

1.5 Types of Variables

In Section 1.4, we learned that a variable is a characteristic under investigation that assumes different values for different elements. The incomes of families, heights of persons, gross sales of companies, prices of college textbooks, makes of cars owned by families, number of accidents, and status (freshman, sophomore, junior, or senior) of students enrolled at a university are a few examples of variables.

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

1.5.1 Quantitative Variables

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

Definition

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

Incomes, heights, gross sales, prices of homes, number of cars owned, and number of accidents are examples of quantitative variables because each of them can be expressed numerically. For instance, the income of a family may be $41,520.75 per year, the gross sales for a company may be $567 million for the past year, and so forth. Such quantitative variables may be classified as either *discrete variables* or *continuous variables.*

Discrete Variables

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

Definition

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

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

Continuous Variables

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

Definition

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

The time taken to complete an examination is an example of a continuous variable because it can assume any value, let us say, between 30 and 60 minutes. The time taken may be 42.6 minutes, 42.67 minutes, or 42.674 minutes. (Theoretically, we can measure time as precisely as we want.) Similarly, the height of a person can be measured to the tenth of an inch or to the hundredth of an inch. However, neither time nor height can be counted in a discrete fashion. A few other examples of continuous variables are weights of people, amount of soda in a 12-ounce can (note that a can does not contain exactly 12 ounces of soda), and yield of potatoes (in pounds) per acre. Note that any variable that involves money is considered a continuous variable.

1.5.2 Qualitative or Categorical Variables

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

Definition

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

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

Figure 1.2 illustrates the types of variables.

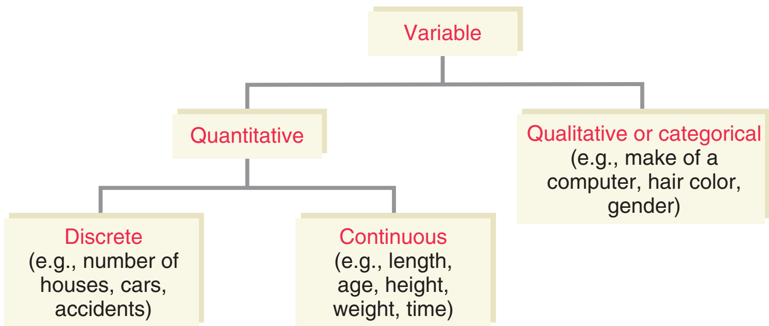


Figure 1.2 Types of variables.

EXERCISES

CONCEPTS AND PROCEDURES

1.13 Explain the meaning of the following terms.

a. Quantitative variable
b. Qualitative variable
c. Discrete variable
d. Continuous variable
e. Quantitative data
f. Qualitative data

APPLICATIONS

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

a. Number of persons in a family
b. Color of cars
c. Marital status of people
d. Length of a frog's jump
e. Number of errors in a person's credit report

1.15 Indicate which of the following variables are quantitative and which are qualitative.
   a. Number of typographical errors in newspapers
   b. Monthly TV cable bills
   c. Spring break locations favored by college students
   d. Number of cars owned by families
   e. Lottery revenues of states

1.16 Classify the quantitative variables in Exercise 1.14 as discrete or continuous.

1.17 Classify the quantitative variables in Exercise 1.15 as discrete or continuous.

1.6 Cross-Section Versus Time-Series Data

Based on the time over which they are collected, data can be classified as either cross-section or time-series data.

1.6.1 Cross-Section Data

Cross-section data contain information on different elements of a population or sample for the *same* period of time. The information on incomes of 100 families for 2005 is an example of cross-section data. All examples of data already presented in this chapter have been cross-section data.

Definition

Cross-Section Data Data collected on different elements at the same point in time or for the same period of time are called *cross-section data*.

Table 1.2 shows the leading daytime talk shows based on the average daily number of viewers from September 2003 to July 2004. Because this table presents data on the average daily viewers for seven shows for the same period (September 2003 to July 2004), it is an example of cross-section data.

Table 1.2 Leading Daytime Talk Shows

Talk Show	Average Daily Viewers (millions)
The Oprah Winfrey Show	8.6
Dr. Phil	6.5
Live With Regis and Kelly	4.7
Maury	4.1
Jerry Springer	3.3
Montel Williams	3.2
Ellen	2.2

Source: Neilsen Media Research; Syndicated Network Television Association. Data appeared in *The New York Times,* August 23, 2004.

1.6.2 Time-Series Data

Time-series data contain information on the same element for *different* periods of time. Information on U.S. exports for the years 1983 to 2005 is an example of time-series data.

Definition

Time-Series Data Data collected on the same element for the same variable at different points in time or for different periods of time are called *time-series data.*

The data given in Table 1.3 are an example of time-series data. This table lists the number of collisions between wildlife (mostly birds) and civilian aircraft that were reported for the years 1990, 1995, 2000, and 2002.

Table 1.3 **Number of Collisions Between Wildlife and Civilian Aircraft**

Year	Number of Collisions
1990	1990
1995	2775
2000	6323
2002	6556

Source: U.S. Air Force, Federal Aviation Administration. *USA Today,* October 15, 2003.

1.7 Sources of Data

The availability of accurate and appropriate data is essential for deriving reliable results.[3] Data may be obtained from internal sources, external sources, or surveys and experiments.

Many times data come from *internal sources*, such as a company's own personnel files or accounting records. For example, a company that wants to forecast the future sales of its product may use the data of past periods from its own records. For most studies, however, all the data that are needed are not usually available from internal sources. In such cases, one may have to depend on outside sources to obtain data. These sources are called *external sources.* For instance, the *Statistical Abstract of the United States* (published annually), which contains various kinds of data on the United States, is an external source of data.

A large number of government and private publications can be used as external sources of data. The following is a list of some of the government publications.

1. *Statistical Abstract of the United States*
2. *Employment and Earnings*
3. *Handbook of Labor Statistics*
4. *Source Book of Criminal Justice Statistics*
5. *Economic Report of the President*
6. *County & City Data Book*
7. *State & Metropolitan Area Data Book*
8. *Digest of Education Statistics*
9. *Health United States*
10. *Agricultural Statistics*

Most of the data contained in these books can be accessed on Internet sites such as www.census.gov (Census Bureau), www.bls.gov (Bureau of Labor Statistics), www.ojp.usdoj.gov/bjs (Office of Justice Program, U.S. Department of Justice, Bureau of Justice Statistics), www.os.dhhs.gov (United States Department of Health and Human Services), and www.usda.gov/nass/pubs/agstats.htm (U.S. Department of Agriculture, Agricultural Statistics).

[3]Sources of data are discussed in more detail in Appendix A.

Besides these government publications, a large number of private publications (e.g., *Standard & Poors' Security Owner's Stock Guide* and *World Almanac and Book of Facts*) and periodicals (e.g., *The Wall Street Journal*, *USA TODAY*, *Fortune*, *Forbes*, and *Business Week*) can be used as external data sources.

Sometimes the needed data may not be available from either internal or external sources. In such cases, the investigator may have to conduct a survey or experiment to obtain the required data. Appendix A discusses surveys and experiments in detail.

EXERCISES

CONCEPTS AND PROCEDURES

1.18 Explain the difference between cross-section and time-series data. Give an example of each of these two types of data.

1.19 Briefly describe internal and external sources of data.

APPLICATIONS

1.20 Classify the following as cross-section or time-series data.
- **a.** Liquor bills of a family for each month of 2005
- **b.** Number of armed robberies each year in Dallas from 1993 to 2005
- **c.** Number of homicides in 40 cities during 2005
- **d.** Gross sales of 200 ice cream parlors in July 2005

1.21 Classify the following as cross-section or time-series data.
- **a.** Average prices of houses in 100 cities
- **b.** Salaries of 50 employees
- **c.** Number of cars sold each year by General Motors from 1980 to 2005
- **d.** Number of employees employed by a company each year from 1985 to 2005

1.8 Summation Notation

Sometimes mathematical notation helps express a mathematical relationship concisely. This section describes the **summation notation** that is used to denote the sum of values.

Suppose a sample consists of five books and the prices of these five books are \$75, \$80, \$35, \$97, and \$88. The variable *price of a book* can be denoted by x. The prices of the five books can be written as follows:

$$\text{Price of the first book} = \underset{\substack{\uparrow \\ \text{Subscript of } x \text{ denotes the} \\ \text{number of the book}}}{x_1} = \$75$$

Similarly,

$$\begin{aligned} \text{Price of the second book} &= x_2 = \$80 \\ \text{Price of the third book} &= x_3 = \$35 \\ \text{Price of the fourth book} &= x_4 = \$97 \\ \text{Price of the fifth book} &= x_5 = \$88 \end{aligned}$$

In this notation, x represents the price, and the subscript denotes a particular book.

Now, suppose we want to add the prices of all five books. We have

$$x_1 + x_2 + x_3 + x_4 + x_5 = 75 + 80 + 35 + 97 + 88 = \$375$$

The uppercase Greek letter Σ (pronounced *sigma*) is used to denote the sum of all values. Using Σ notation, we can write the foregoing sum as follows:

$$\Sigma x = x_1 + x_2 + x_3 + x_4 + x_5 = \$375$$

The notation Σx in this expression represents the sum of all the values of x and is read as "sigma x" or "sum of all values of x."

EXAMPLE 1–1

Using summation notation: one variable.

Annual salaries (in thousands of dollars) of four workers are 75, 42, 125, and 61. Find

(a) Σx **(b)** $(\Sigma x)^2$ **(c)** Σx^2

Solution Let x_1, x_2, x_3, and x_4 be the annual salaries (in thousands of dollars) of the first, second, third, and fourth worker, respectively. Then,

$$x_1 = 75, \quad x_2 = 42, \quad x_3 = 125, \quad \text{and} \quad x_4 = 61$$

(a) $\Sigma x = x_1 + x_2 + x_3 + x_4 = 75 + 42 + 125 + 61 = 303 = \mathbf{\$303{,}000}$

(b) Note that $(\Sigma x)^2$ is the square of the sum of all x values. Thus,

$$(\Sigma x)^2 = (303)^2 = \mathbf{91{,}809}$$

(c) The expression Σx^2 is the sum of the squares of x values. To calculate Σx^2, we first square each of the x values and then sum these squared values. Thus

$$\begin{aligned}\Sigma x^2 &= (75)^2 + (42)^2 + (125)^2 + (61)^2 \\ &= 5625 + 1764 + 15{,}625 + 3721 = \mathbf{26{,}735}\end{aligned}$$

EXAMPLE 1–2

Using summation notation: two variables.

The following table lists four pairs of m and f values:

m	12	15	20	30
f	5	9	10	16

Compute the following:

(a) Σm **(b)** Σf^2 **(c)** Σmf **(d)** $\Sigma m^2 f$

Solution We can write

$$\begin{array}{llll} m_1 = 12 & m_2 = 15 & m_3 = 20 & m_4 = 30 \\ f_1 = 5 & f_2 = 9 & f_3 = 10 & f_4 = 16 \end{array}$$

(a) $\Sigma m = 12 + 15 + 20 + 30 = \mathbf{77}$

(b) $\Sigma f^2 = (5)^2 + (9)^2 + (10)^2 + (16)^2 = 25 + 81 + 100 + 256 = \mathbf{462}$

(c) To compute Σmf, we multiply the corresponding values of m and f and then add the products as follows:

$$\begin{aligned}\Sigma mf &= m_1 f_1 + m_2 f_2 + m_3 f_3 + m_4 f_4 \\ &= 12(5) + 15(9) + 20(10) + 30(16) = \mathbf{875}\end{aligned}$$

(d) To calculate $\Sigma m^2 f$, we square each m value, then multiply the corresponding m^2 and f values, and add the products. Thus,

$$\begin{aligned}\Sigma m^2 f &= (m_1)^2 f_1 + (m_2)^2 f_2 + (m_3)^2 f_3 + (m_4)^2 f_4 \\ &= (12)^2(5) + (15)^2(9) + (20)^2(10) + (30)^2(16) = \mathbf{21{,}145}\end{aligned}$$

The calculations done in parts (a) through (d) to find the values of Σm, Σf^2, Σmf, and $\Sigma m^2 f$ can be performed in tabular form, as shown in Table 1.4.

Table 1.4

m	f	f^2	mf	m^2f
12	5	$5 \times 5 = 25$	$12 \times 5 = 60$	$12 \times 12 \times 5 = 720$
15	9	$9 \times 9 = 81$	$15 \times 9 = 135$	$15 \times 15 \times 9 = 2025$
20	10	$10 \times 10 = 100$	$20 \times 10 = 200$	$20 \times 20 \times 10 = 4000$
30	16	$16 \times 16 = 256$	$30 \times 16 = 480$	$30 \times 30 \times 16 = 14{,}400$
$\Sigma m = 77$	$\Sigma f = 40$	$\Sigma f^2 = 462$	$\Sigma mf = 875$	$\Sigma m^2f = 21{,}145$

The columns of Table 1.4 can be explained as follows.

1. The first column lists the values of m. The sum of these values gives $\Sigma m = 77$.
2. The second column lists the values of f. The sum of this column gives $\Sigma f = 40$.
3. The third column lists the squares of the f values. For example, the first value, 25, is the square of 5. The sum of the values in this column gives $\Sigma f^2 = 462$.
4. The fourth column records products of the corresponding m and f values. For example, the first value, 60, in this column is obtained by multiplying 12 by 5. The sum of the values in this column gives $\Sigma mf = 875$.
5. Next, the m values are squared and multiplied by the corresponding f values. The resulting products, denoted by m^2f, are recorded in the fifth column. For example, the first value, 720, is obtained by squaring 12 and multiplying this result by 5. The sum of the values in this column gives $\Sigma m^2f = 21{,}145$. ■

EXERCISES

CONCEPTS AND PROCEDURES

1.22 The following table lists five pairs of m and f values.

m	5	10	17	20	25
f	12	8	6	16	4

Compute the value of each of the following:

a. Σm **b.** Σf^2 **c.** Σmf **d.** Σm^2f

1.23 The following table lists six pairs of m and f values.

m	3	6	25	12	15	18
f	16	11	16	8	4	14

Calculate the value of each of the following:

a. Σf **b.** Σm^2 **c.** Σmf **d.** Σm^2f

1.24 The following table lists five pairs of x and y values.

x	15	22	11	8	5
y	10	12	14	9	18

Compute

a. Σx **b.** Σy **c.** Σxy **d.** Σx^2 **e.** Σy^2

1.25 The following table lists six pairs of x and y values.

x	4	18	25	9	12	20
y	12	5	14	7	12	8

Compute

a. Σx **b.** Σy **c.** Σxy **d.** Σx^2 **e.** Σy^2

APPLICATIONS

1.26 Six adults spent \$20, \$14, \$57, \$23, \$7, and \$102 on lottery tickets last month. Let y denote last month's lottery ticket expenses for an adult. Find

a. Σy **b.** $(\Sigma y)^2$ **c.** Σy^2

1.27 The phone bills for January 2006 for four families were \$83, \$205, \$87, and \$154. Let y be the amount of the January 2006 phone bill for a family. Find

a. Σy **b.** $(\Sigma y)^2$ **c.** Σy^2

1.28 Prices (in thousands of dollars) of five new cars are 28, 35, 29, 54, and 18. Let x be the price of a new car in this sample. Find

a. Σx **b.** $(\Sigma x)^2$ **c.** Σx^2

1.29 The number of students (rounded to the nearest thousand) currently enrolled at seven universities are 7, 39, 21, 16, 3, 43, and 19. Let x be the number of students currently enrolled at a university. Find

a. Σx **b.** $(\Sigma x)^2$ **c.** Σx^2

USES AND MISUSES... SPEAKING THE LANGUAGE OF STATISTICS

Have you ever heard the statistic "the average American family has 2.1 children"? What is wrong with this statement, and how do we fix it? How about: "In a representative sample of 10 American families, one can expect there to be 21 children." The statement is wordy but more accurate. Why do we care?

Statisticians pay close attention to definitions because, without them, calculations would be impossible to make and interpretations of the data would be meaningless. Often, when you read statistics reported in the newspaper, the journalist or editor sometimes chooses to describe the results in a way that is easier to understand but that distorts the actual statistical result.

Let's pick apart our example. The word *average* has a very specific meaning in probability (Chapters 4 and 5). The intended meaning of the word here really is *typical.* The adjective *American* helps us define the population. The Census Bureau defines *family* as "a group of two people or more (one of whom is the householder) related by birth, marriage, or adoption and residing together; all such people (including related subfamily members) are considered as members of one family." It defines *children* as "all persons under 18 years, excluding people who maintain households, families, or subfamilies as a reference person or spouse." We understand implicitly that a family cannot have a fractional number of children, so we accept that this discrete variable takes on the properties of a continuous variable when we are talking about the characteristics of a large population. How large does the population need to be before we can derive continuous variables from discrete variables? The answer comes in the chapters that follow.

The moral of the story is that whenever you read a statistical result, be sure that you understand the definitions of the terms used to describe the result and relate those terms to the definitions that you already know. In some cases *year* is a categorical variable, in others it is a discrete variable, in others a continuous variable. Many surveys will report that "respondents feel better, the same, or worse" about a particular subject. Although *better*, *same*, and *worse* have a natural order to them, they do not have numerical values.

Glossary

Census A survey that includes all members of the population.

Continuous variable A (quantitative) variable that can assume any numerical value over a certain interval or intervals.

Cross-section data Data collected on different elements at the same point in time or for the same period of time.

Data or **data set** Collection of observations or measurements on a variable.

Descriptive statistics Collection of methods for organizing, displaying, and describing data using tables, graphs, and summary measures.

Discrete variable A (quantitative) variable whose values are countable.

Element or **member** A specific subject or object included in a sample or population.

Inferential statistics Collection of methods that help make decisions about a population based on sample results.

Observation or **measurement** The value of a variable for an element.

Population or **target population** The collection of all elements whose characteristics are being studied.

Qualitative or **categorical data** Data generated by a qualitative variable.

Qualitative or **categorical variable** A variable that cannot assume numerical values but is classified into two or more categories.

Quantitative data Data generated by a quantitative variable.

Quantitative variable A variable that can be measured numerically.

Random sample A sample drawn in such a way that each element of the population has some chance of being included in the sample.

Representative sample A sample that contains the same characteristics as the corresponding population.

Sample A portion of the population of interest.

Sample survey A survey that includes elements of a sample.

Simple random sampling If all samples of the same size selected from a population have the same chance of being selected, it is called simple random sampling. Such a sample is called a simple random sample.

Statistics Group of methods used to collect, analyze, present, and interpret data and to make decisions.

Survey Collection of data on the elements of a population or sample.

Time-series data Data that give the values of the same variable for the same element at different points in time or for different periods of time.

Variable A characteristic under study or investigation that assumes different values for different elements.

Supplementary Exercises

1.30 The following table gives the average attendance at interleague Major League baseball games for the 1999 to 2004 seasons.

Season	Average Attendance at Interleague Games
1999	33,352
2000	33,212
2001	33,692
2002	31,921
2003	30,894
2004	32,976

Source: USA TODAY, July 8, 2004.

Describe the meaning of a variable, a measurement, and a data set with reference to this table.

1.31 The following table lists the number of Americans who took cruises during 1995 to 2004.

Year	Total Passengers (millions)
1995	4.4
1996	4.7
1997	5.1
1998	5.4
1999	5.9
2000	6.9
2001	6.9
2002	7.6
2003	8.2
2004	9.0

Source: Cruise Lines International Association. *USA TODAY,* February 4, 2005.

Describe the meaning of a variable, a measurement, and a data set with reference to this table.

1.32 Refer to Exercises 1.30 and 1.31. Classify these data sets as either cross-section or time-series.

1.33 Indicate whether each of the following examples refers to a population or to a sample.

a. A group of 25 patients selected to test a new drug
b. Total items produced on a machine for each year from 1995 to 2005
c. Yearly expenditures on clothes for 50 persons
d. Number of houses sold by each of the 10 employees of a real estate agency during 2005

1.34 Indicate whether each of the following examples refers to a population or to a sample.

- **a.** Salaries of CEOs of all companies in New York City
- **b.** Allowances of 1500 sixth-graders selected from Ohio
- **c.** Gross sales for 2005 of four fast-food chains
- **d.** Annual incomes of all 33 employees of a restaurant

1.35 State which of the following is an example of sampling with replacement and which is an example of sampling without replacement.

- **a.** Selecting 10 patients out of 100 to test a new drug
- **b.** Selecting one professor to be a member of the university senate and then selecting one professor from the same group to be a member of the curriculum committee

1.36 State which of the following is an example of sampling with replacement and which is an example of sampling without replacement.

- **a.** Selecting seven cities to market a new deodorant
- **b.** Selecting a high school teacher to drive students to a lecture in March, then selecting a teacher from the same group to chaperone a dance in April

1.37 The number of shoe pairs owned by six women are 8, 14, 3, 7, 10, and 5. Let x denote the number of shoe pairs owned by a woman. Find

a. Σx **b.** $(\Sigma x)^2$ **c.** Σx^2

1.38 The number of restaurants in each of five small towns is 4, 12, 8, 10, and 5. Let y denote the number of restaurants in a small town. Find

a. Σy **b.** $(\Sigma y)^2$ **c.** Σy^2

1.39 The following table lists five pairs of m and f values.

m	3	16	11	9	20
f	7	32	17	12	34

Compute the value of each of the following:

a. Σm **b.** Σf^2 **c.** Σmf **d.** $\Sigma m^2 f$ **e.** Σm^2

1.40 The following table lists six pairs of x and y values.

x	7	11	8	4	14	28
y	5	15	7	10	9	19

Compute the value of each of the following:

a. Σy **b.** Σx^2 **c.** Σxy **d.** $\Sigma x^2 y$ **e.** Σy^2

Self-Review Test

1. A population in statistics means a collection of all

- **a.** men and women
- **b.** subjects or objects of interest
- **c.** people living in a country

2. A sample in statistics means a portion of the

- **a.** people selected from the population of a country
- **b.** people selected from the population of an area
- **c.** population of interest

3. Indicate which of the following is an example of a sample with replacement and which is a sample without replacement.

- **a.** Five friends go to a livery stable and select five horses to ride (each friend must choose a different horse).
- **b.** A box contains five balls of different colors. A ball is drawn from this box, its color is recorded, and it is put back into the box before the next ball is drawn. This experiment is repeated 12 times.

4. Indicate which of the following variables are quantitative and which are qualitative. Classify the quantitative variables as discrete or continuous.

a. Women's favorite TV programs
b. Salaries of football players
c. Number of pets owned by families
d. Favorite breed of dog for each of 20 children

5. The following table lists the total career yards rushed by each of the top ten rushers in the National Football League as of February 4, 2005.

Player	Yards
Emmitt Smith	18,355
Walter Payton	16,726
Barry Sanders	15,269
Curtis Martin	13,336
Jerome Bettis	13,294
Eric Dickerson	13,259
Tony Dorsett	12,739
Jim Brown	12,312
Marcus Allen	12,243
Franco Harris	12,120

Source: USA TODAY, February 4, 2005.

Explain the meaning of a member, a variable, a measurement, and a data set with reference to this table.

6. The number of credit cards possessed by five couples are 2, 5, 3, 12, and 7. Let x be the number of credit cards possessed by a couple. Find:

a. Σx **b.** $(\Sigma x)^2$ **c.** Σx^2

7. The following table lists five pairs of m and f values.

m	3	6	9	12	15
f	15	25	40	20	12

Calculate

a. Σm **b.** Σf **c.** Σm^2 **d.** Σmf **e.** $\Sigma m^2 f$ **f.** Σf^2

Mini-Project

■ MINI-PROJECT 1–1

In this mini-project, you are going to obtain a data set of interest to you that you will use for mini-projects in other chapters throughout the course. The data set should contain at least one qualitative variable and one quantitative variable, although having two of each will be necessary in some cases. Ask your instructor how many variables you should have. A good size data set to work with should contain somewhere between 50 and 100 observations.

Here are some examples of the procedures to use to obtain data:

1. Take a random sample of used cars and collect data on them. You may use Web sites like Cars.com, AutoTrader.com, and so forth. Quantitative variables may include the price, mileage, and age of a car. Categorical variables may include the model, drive train (front wheel, rear wheel, and so forth), and type (compact, SUV, minivan, and so forth). You can concentrate on your favorite type of car, or look at a variety of types.

2. Examine the real estate ads in your local newspaper or online and obtain information on rental properties or houses for sale.

3. Use an almanac or go to a government Web site, such as www.census.gov or www.cdc.gov to obtain information for each state. Quantitative variables may include income, birth and death rates, cancer incidence, and the proportion of people living below the poverty level. Categorical variables may include things like the region of the country where each state is located and which party won the state governorship in the last election. You can also collect this information on a worldwide level and use the continent or world region as a categorical variable.

4. Take a random sample of students and ask them questions such as:

- How much money did you spend on books last semester?
- How many credit hours did you take?
- What is your major?

5. If you are a sports fan, you can use an almanac or sports Web site to obtain statistics on a random sample of athletes. You can look at sport-specific statistics such as home runs, runs batted in, position, left-handed/right-handed, and so forth in baseball, or you could collect information to compare different sports by gathering information on salary, career length, weight, and so forth.

Once you have collected the information, write a brief report that includes answers to the following tasks/questions:

a. Describe the variables that you have collected information on.
b. Describe a reasonable target population for the sample you used.
c. Is your sample a random sample from this target population?
d. Do you feel that your sample is representative of this population?
e. Is this an example of sampling with or without replacement?
f. For each quantitative variable, state whether it is continuous or discrete.
g. Describe the meaning of an element, a variable, and a measurement for this data set.
h. Describe any problems you faced in collecting these data.
i. Were any of the data values unusable? If yes, explain why.

Your instructor will probably want to see a copy of the data you collected. If you are using statistical software in the class, enter the data into that software and submit a copy of the data file. If you are using a handheld technology calculator, such as a graphing calculator, you will probably have to print out a hard copy version of the data set. Save this data set for projects in future chapters.

TECHNOLOGY INSTRUCTION — Entering and Saving Data

Whenever you want to analyze and interpret some data, you need to enter those data in some technology, proofread it, and revise it. If you will be using the data entered into a technology again at a later date, you need to save these data into your technology so that you can retrieve them later.

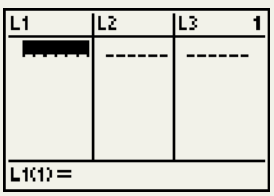
TI-84

1. On the TI-84, variables are referred to as lists.

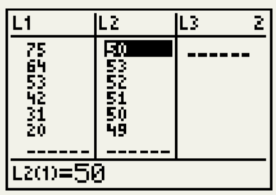
2. In order to enter data into the TI-84 calculator, you first need to decide if you will want to save the data for later use or you want to use it only in the immediate future.

3. If you will be using these data only in the immediate future, select **STAT>EDIT>SetUp-Editor**, and then press **Enter**. This will set up the editor to use "scratch" lists **L1, L2, L3, L4, L5,** and **L6** (see **Screen 1.1**). Now select **STAT>EDIT>Edit** and start typing your numeric data into the column or columns and press **Enter** after each entry (see **Screen 1.2**). Note: The TI-84 will not handle non-numeric data.

Screen 1.1



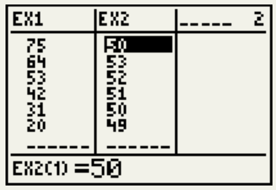
Screen 1.2



4. If you will be using your data at a later date, it is better to give the variables names so that you do not have to reenter the data. Select **STAT>EDIT>SetUpEditor,** and then type in the names of your variables separted by commas (see **Screens 1.3 and 1.4**). Names can be 1 to 5 letters long. These letters can be found in green on the TI-84 keyboard. You can use the green **ALPHA** key with each letter, or press **A-LOCK** (**2nd>ALPHA**) while you are typing the name. To turn off **A-LOCK**, press **ALPHA**. Note that here we have used the names EX1 and EX2. You can use any names such as Price, Brand, and so forth.

Screen 1.3

Screen 1.4



5. You can use the arrow keys to move around and go back to a cell to edit its contents. When editing values, you will need to press **Enter** for the changes to take effect.
6. **SetUpEditor** determines what lists are displayed in the editor. Changing what **SetUpEditor** displays does not delete any lists. Your lists remain in storage when the calculator is turned off.

MINITAB

1. Start MINITAB by clicking the MINITAB icon. You will see a session window as well as a worksheet, similar to a spreadsheet, where you will enter your data (see **Screen 1.5**).

Screen 1.5

↓	C1	C2	C3-T
	year	sales	employee
1	2001	35	J. Smith
2	2001	38	A. Jones
3	2002	50	J. Smith
4	2002	48	A. Jones

2. Use the mouse or the arrow keys to select a cell in the MINITAB worksheet where you want to start entering your data. Data can be numeric, text, or date/time, but you can enter only one type of data into a given column since the columns correspond to variables. The rectangles in the Worksheet are called cells, and the cells are organized into columns such as **C1, C2, . . .**, and so forth. Each column has rows numbered **1, 2, . . .**, and so forth. Note that if a column contains text data as column C3 in Screen 1.5, "**-T**" will be added to the column heading.
3. The blank row between the column labels and row 1 is for variable names.
4. You can change whether you are typing the data across in rows or down in columns by clicking the direction arrow at the top left of the worksheet (see **Screen 1.5**).
5. If you want to type a table in a given array of cells, then select that set of cells and begin typing.
6. If you need to revise an entry, go to that cell with the mouse or the arrow keys and begin typing. Press **Enter** to put the revised entry into the cell.

7. When you are done, select **File>Save Current Worksheet As** to save your work for the first time as a file on your computer. Note that MINITAB will automatically assign the file extension *.mtw* to your work after you choose the filename.

8. Try entering the following data into MINITAB:

January	52	.08
February	48	.06
March	49	.07

Name the columns *Month, Sales, Increase.* Save the typed data as *test.mtw* file.

9. To retrieve the file, select **File>Open** and select the *test.mtw* file.

10. If you are already in MINITAB and you want to start a new worksheet, select **File>New** and choose **Worksheet**.

Excel

1. Start Excel.

2. Use the mouse or the arrow keys to select the cell in the spreadsheet where you want to start entering your data. Data can be numeric or text. The rectangles are called cells, and the cells are collectively known as a spreadsheet.

3. You can format your data by selecting the cells that you want to format, then selecting **Format>Cells**, and then choosing whether you want to format a number, align text, and so forth. For common formatting tasks, you have icons on the toolbar, such as a dollar sign ($) to format currency, a percent sign (%) to format numbers as percents, and icons representing left, center, and right aligned text to change your alignment.

4. If you need to revise an entry, go to that cell with the mouse or the arrow keys. You can retype the entry or you can edit it. To edit it, select the **Formula Bar** at the top of the screen (which contains the contents of the cell) and use the arrow keys and the backspace key to help you revise the entry, then press **Enter** to put the revised entry into the cell.

5. When you are done, select **File>Save As** to save your work for the first time as a file on your computer. Note that Excel will automatically assign the file extension *.xls* to your work after you choose the filename.

6. Try entering the following data into Excel:

January	52	.08
February	48	.06
March	49	.07

Format it to look like the following.

January	$52.00	8%
February	$48.00	6%
March	$49.00	7%

Save the result as *test.xls* file.

7. To retrieve this file, select **File>Open** and select the *test.xls* file.

Screen 1.6 shows the data of Screen 1.5 as entered in Excel.

	A	B	C
1	year	sales	employee
2	2001	35	J. Smith
3	2001	38	A. Jones
4	2002	50	J. Smith
5	2002	48	A. Jones
6			

Screen 1.6

TECHNOLOGY ASSIGNMENTS

TA1.1 The following table gives the names, hours worked, and salary for the past week for five workers.

Name	Hours Worked	Salary
John	42	$725
Shannon	33	1583
Kathy	28	1255
David	47	1090
Steve	40	820

a. Enter these data into the spreadsheet. Save the data file as WORKER. Exit the session or program. Then restart the program or software and retrieve the file WORKER.

b. Print a hard copy of the spreadsheet containing data you entered.

TA1.2 Refer to data on 2004 profits of seven companies given in Table 1.1 on page 9 Enter those data into the spreadsheet and save this file as PROFITS.