CHAPTER 14

Sampling and Simulation

Objectives

After completing this chapter, you should be able to

1. Demonstrate a knowledge of the four basic sampling methods.
2. Recognize faulty questions on a survey and other factors that can bias responses.
3. Solve problems, using simulation techniques.

Outline

Statistics Today

The Monty Hall Problem

On the game show *Let's Make A Deal,* host Monty Hall gave a contestant a choice of three doors. A valuable prize was behind one door, and nothing was behind the other two doors. When the contestant selected one door, host Monty Hall opened one of the other doors that the contestant didn't select and that had no prize behind it. (Monty Hall knew in advance which door had the prize.) Then he asked the contestant if he or she wanted to change doors or keep the one that the contestant originally selected. Now the question is, Should the contestant switch doors, or does it really matter? This chapter will show you how you can solve this problem by simulation. For the answer, see Statistics Today—Revisited.

14–1 Introduction

Most people have heard of Gallup, Harris, and Nielsen. These and other pollsters gather information about the habits and opinions of the U.S. people. Such survey firms, and the U.S. Census Bureau, gather information by selecting samples from well-defined populations. Recall from Chapter 1 that the subjects in the sample should be a subgroup of the subjects in the population. Sampling methods often use what are called *random numbers* to select samples.

Since many statistical studies use surveys and questionnaires, some information about these is presented in Section 14–3.

Random numbers are also used in *simulation techniques.* Instead of studying a real-life situation, which may be costly or dangerous, researchers create a similar situation in a laboratory or with a computer. Then, by studying the simulated situation, researchers can gain the necessary information about the real-life situation in a less expensive or safer manner. This chapter will explain some common methods used to obtain samples as well as the techniques used in simulations.

14–2

Common Sampling Techniques

Objective 1

Demonstrate a knowledge of the four basic sampling methods.

In Chapter 1, a *population* was defined as all subjects (human or otherwise) under study. Since some populations can be very large, researchers cannot use every single subject, so a sample must be selected. A *sample* is a subgroup of the population. Any subgroup of the population, technically speaking, can be called a sample. However, for researchers to make valid inferences about population characteristics, the sample must be random.

> For a sample to be a **random sample,** every member of the population must have an equal chance of being selected.

When a sample is chosen at random from a population, it is said to be an **unbiased sample.** That is, the sample, for the most part, is representative of the population. Conversely, if a sample is selected incorrectly, it may be a biased sample. Samples are said to be **biased samples** when some type of systematic error has been made in the selection of the subjects.

A sample is used to get information about a population for several reasons:

1. *It saves the researcher time and money.*
2. *It enables the researcher to get information that he or she might not be able to obtain otherwise.* For example, if a person's blood is to be analyzed for cholesterol, a researcher cannot analyze every single drop of blood without killing the person. Or if the breaking strength of cables is to be determined, a researcher cannot test to destruction every cable manufactured, since the company would not have any cables left to sell.
3. *It enables the researcher to get more detailed information about a particular subject.* If only a few people are surveyed, the researcher can conduct in-depth interviews by spending more time with each person, thus getting more information about the subject. This is not to say that the smaller the sample, the better; in fact, the opposite is true. In general, larger samples—if correct sampling techniques are used—give more reliable information about the population.

It would be ideal if the sample were a perfect miniature of the population in all characteristics. This ideal, however, is impossible to achieve, because there are so many human traits (height, weight, IQ, etc.). The best that can be done is to select a sample that will be representative with respect to *some* characteristics, preferably those pertaining to the study. For example, if one-half of the population subjects are female, then approximately one-half of the sample subjects should be female. Likewise, other characteristics, such as age, socioeconomic status, and IQ, should be represented proportionately. To obtain unbiased samples, statisticians have developed several basic sampling methods. The most common methods are *random, systematic, stratified,* and *cluster sampling*. Each method will be explained in detail in this section.

In addition to the basic methods, there are other methods used to obtain samples. Some of these methods are also explained in this section.

Random Sampling

A random sample is obtained by using methods such as random numbers, which can be generated from calculators, computers, or tables. In *random sampling,* the basic requirement is that for a sample of size *n,* all possible samples of this size must have an equal chance of being selected from the population. But before the correct method of obtaining a random sample is explained, several incorrect methods commonly used by various researchers and agencies to gain information are discussed.

One incorrect method commonly used is to ask "the person on the street." News reporters use this technique quite often. Selecting people haphazardly on the street does not meet the requirement for simple random sampling, since not all possible samples of a specific size have an equal chance of being selected. Many people will be at home or at work when the interview is being conducted and therefore do not have a chance of being selected.

Another incorrect technique is to ask a question by either radio or television and have the listeners or viewers call the station to give their responses or opinions. Again, this sample is not random, since only those who feel strongly for or against the issue may respond and people may not have heard or seen the program. A third erroneous method is to ask people to respond by mail. Again, only those who are concerned and who have the time are likely to respond.

These methods do not meet the requirement of random sampling, since not all possible samples of a specific size have an equal chance of being selected. To meet this requirement, researchers can use one of two methods. The first method is to number each element of the population and then place the numbers on cards. Place the cards in a hat or fishbowl, mix them, and then select the sample by drawing the cards. When using this procedure, researchers must ensure that the numbers are well mixed. On occasion, when this procedure is used, the numbers are not mixed well, and the numbers chosen for the sample are those that were placed in the bowl last.

The second and preferred way of selecting a random sample is to use random numbers. Figure 14–1 shows a table of two-digit random numbers generated by a computer. A more detailed table of random numbers is found in Table D of Appendix C.

The theory behind random numbers is that each digit, 0 through 9, has an equal probability of occurring. That is, in every sequence of 10 digits, each digit has a probability of $\frac{1}{10}$ of occurring. This does not mean that in every sequence of 10 digits, one will find each digit. Rather, it means that on the average, each digit will occur once. For example, the digit 2 may occur 3 times in a sequence of 10 digits, but in later sequences, it may not occur at all, thus averaging to a probability of $\frac{1}{10}$.

To obtain a sample by using random numbers, number the elements of the population sequentially and then select each person by using random numbers. This process is shown in Example 14–1.

Random samples can be selected with or without replacement. If the same member of the population cannot be used more than once in the study, then the sample is selected without replacement. That is, once a random number is selected, it cannot be used later.

Figure 14–1

Table of 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

Note: In the explanations and examples of the sampling procedures, a small population will be used, and small samples will be selected from this population. Small populations are used for illustrative purposes only, because the entire population could be included with little difficulty. In real life, however, researchers must usually sample from very large populations, using the procedures shown in this chapter.

Example 14–1

Suppose a researcher wants to produce a television show featuring in-depth interviews with state governors on the subject of capital punishment. Because of time constraints, the 60-minute program will have room for only 10 governors. The researcher wishes to select the governors at random. Select a random sample of 10 states from 50.

Note: This answer is not unique.

Solution

Step 1 Number each state from 1 to 50, as shown. In this case, they are numbered alphabetically.

01. Alabama	14. Indiana	27. Nebraska	40. South Carolina
02. Alaska	15. Iowa	28. Nevada	41. South Dakota
03. Arizona	16. Kansas	29. New Hampshire	42. Tennessee
04. Arkansas	17. Kentucky	30. New Jersey	43. Texas
05. California	18. Louisiana	31. New Mexico	44. Utah
06. Colorado	19. Maine	32. New York	45. Vermont
07. Connecticut	20. Maryland	33. North Carolina	46. Virginia
08. Delaware	21. Massachusetts	34. North Dakota	47. Washington
09. Florida	22. Michigan	35. Ohio	48. West Virginia
10. Georgia	23. Minnesota	36. Oklahoma	49. Wisconsin
11. Hawaii	24. Mississippi	37. Oregon	50. Wyoming
12. Idaho	25. Missouri	38. Pennsylvania	
13. Illinois	26. Montana	39. Rhode Island	

Step 2 Using the random numbers shown in Figure 14–1, find a starting point. To find a starting point, one generally closes one's eyes and places one's finger anywhere on the table. In this case, the first number selected was 27 in the fourth column. Going down the column and continuing on to the next column, select the first 10 numbers. They are 27, 95, 27, 73, 60, 43, 56, 34, 93, and 06. See Figure 14–2. (Note that 06 represents 6.)

Figure 14–2

Selecting a Starting Point and 10 Numbers from the Random Number Table

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	*Start here	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

Now, refer to the list of states and identify the state corresponding to each number. The sample consists of the following states:

27	Nebraska	43	Texas
95		56	
27	Nebraska	34	North Dakota
73		93	
60		06	Colorado

Step 3 Since the numbers 95, 73, 60, 56, and 93 are too large, they are disregarded. And since 27 appears twice, it is also disregarded the second time. Now, one must select six more random numbers between 1 and 50 and omit duplicates, since this sample will be selected without replacement. Make this selection by continuing down the column and moving over to the next column until a total of 10 numbers is selected. The final 10 numbers are 27, 43, 34, 06, 13, 29, 01, 39, 23, and 35. See Figure 14–3.

Figure 14–3

The Final 10 Numbers Selected

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

These numbers correspond to the following states:

27	Nebraska	29	New Hampshire
43	Texas	01	Alabama
34	North Dakota	39	Rhode Island
06	Colorado	23	Minnesota
13	Illinois	35	Ohio

Thus, the governors of these 10 states will constitute the sample.

Random sampling has one limitation. If the population is extremely large, it is time-consuming to number and select the sample elements. Also, notice that the random numbers in the table are two-digit numbers. If three digits are needed, then the first digit from the next column can be used, as shown in Figure 14–4. Table D in Appendix C gives five-digit random numbers.

Speaking of Statistics

Should We Be Afraid of Lightning?

The National Weather Service collects various types of data about the weather. For example, each year in the United States about 400 million lightning strikes occur. On average, 400 people are struck by lightning, and 85% of those struck are men. About 100 of these people die. The cause of most of these deaths is not burns, even though temperatures as high as 54,000°F are reached, but heart attacks. The lightning strike short-circuits the body's autonomic nervous system, causing the heart to stop beating. In some instances, the heart will restart on its own. In other cases, the heart victim will need emergency resuscitation.

The most dangerous places to be during a thunderstorm are open fields, golf courses, under trees, and near water, such as a lake or swimming pool. It's best to be inside a building during a thunderstorm although there's no guarantee that the building won't be struck by lightning. Are these statistics descriptive or inferential? Why do you think more men are struck by lightning than women? Should you be afraid of lightning?

Figure 14–4

Method for Selecting Three-Digit 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

Use one column and part of the next column for three digits, that is, 404.

Systematic Sampling

A **systematic sample** is a sample obtained by numbering each element in the population and then selecting every third or fifth or tenth, etc., number from the population to be included in the sample. This is done after the first number is selected at random.

The procedure of systematic sampling is illustrated in Example 14–2.

Example 14–2

Using the population of 50 states in Example 14–1, select a systematic sample of 10 states.

Solution

Step 1 Number the population units as shown in Example 14–1.

Step 2 Since there are 50 states and 10 are to be selected, the rule is to select every fifth state. This rule was determined by dividing 50 by 10, which yields 5.

Step 3 Using the table of random numbers, select the first digit (from 1 to 5) at random. In this case, 4 was selected.

Step 4 Select every fifth number on the list, starting with 4. The numbers include the following:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 . . .

The selected states are as follows:

4	Arkansas	29	New Hampshire
9	Florida	34	North Dakota
14	Indiana	39	Rhode Island
19	Maine	44	Utah
24	Mississippi	49	Wisconsin

The advantage of systematic sampling is the ease of selecting the sample elements. Also, in many cases, a numbered list of the population units may already exist. For example, the manager of a factory may have a list of employees who work for the company, or there may be an in-house telephone directory.

When doing systematic sampling, one must be careful how the items are arranged on the list. For example, if each unit were arranged, say, as

1. Husband
2. Wife
3. Husband
4. Wife

then the selection of the starting number could produce a sample of all males or all females, depending on whether the starting number is even or odd and whether the number to be added is even or odd. As another example, if the list were arranged in order of heights of individuals, one would get a different average from two samples if the first were selected by using a small starting number and the second by using a large starting number.

Stratified Sampling

A **stratified sample** is a sample obtained by dividing the population into subgroups, called *strata,* according to various homogeneous characteristics and then selecting members from each stratum for the sample.

For example, a population may consist of males and females who are smokers or nonsmokers. The researcher will want to include in the sample people from each group—that is, males who smoke, males who do not smoke, females who smoke, and females who do not smoke. To accomplish this selection, the researcher divides the population into four subgroups and then selects a random sample from each subgroup. This method ensures that the sample is representative on the basis of the characteristics of gender and smoking. Of course, it may not be representative on the basis of other characteristics.

Example 14–3

Using the population of 20 students shown in Figure 14–5, select a sample of eight students on the basis of gender (male/female) and grade level (freshman/sophomore) by stratification.

Figure 14–5
Population of Students for Example 14–3

1. Ald, Peter	M	Fr	11. Martin, Janice	F	Fr
2. Brown, Danny	M	So	12. Meloski, Gary	M	Fr
3. Bear, Theresa	F	Fr	13. Oeler, George	M	So
4. Carson, Susan	F	Fr	14. Peters, Michele	F	So
5. Collins, Carolyn	F	Fr	15. Peterson, John	M	Fr
6. Davis, William	M	Fr	16. Smith, Nancy	F	Fr
7. Hogan, Michael	M	Fr	17. Thomas, Jeff	M	So
8. Jones, Lois	F	So	18. Toms, Debbie	F	So
9. Lutz, Harry	M	So	19. Unger, Roberta	F	So
10. Lyons, Larry	M	So	20. Zibert, Mary	F	So

Solution

Step 1 Divide the population into two subgroups, consisting of males and females, as shown in Figure 14–6.

Figure 14–6
Population Divided into Subgroups by Gender

Males			Females		
1. Ald, Peter	M	Fr	1. Bear, Theresa	F	Fr
2. Brown, Danny	M	So	2. Carson, Susan	F	Fr
3. Davis, William	M	Fr	3. Collins, Carolyn	F	Fr
4. Hogan, Michael	M	Fr	4. Jones, Lois	F	So
5. Lutz, Harry	M	So	5. Martin, Janice	F	Fr
6. Lyons, Larry	M	So	6. Peters, Michele	F	So
7. Meloski, Gary	M	Fr	7. Smith, Nancy	F	Fr
8. Oeler, George	M	So	8. Toms, Debbie	F	So
9. Peterson, John	M	Fr	9. Unger, Roberta	F	So
10. Thomas, Jeff	M	So	10. Zibert, Mary	F	So

Step 2 Divide each subgroup further into two groups of freshmen and sophomores, as shown in Figure 14–7.

Figure 14–7
Each Subgroup Divided into Subgroups by Grade Level

Group 1			Group 2		
1. Ald, Peter	M	Fr	1. Bear, Theresa	F	Fr
2. Davis, William	M	Fr	2. Carson, Susan	F	Fr
3. Hogan, Michael	M	Fr	3. Collins, Carolyn	F	Fr
4. Meloski, Gary	M	Fr	4. Martin, Janice	F	Fr
5. Peterson, John	M	Fr	5. Smith, Nancy	F	Fr

Group 3			Group 4		
1. Brown, Danny	M	So	1. Jones, Lois	F	So
2. Lutz, Harry	M	So	2. Peters, Michele	F	So
3. Lyons, Larry	M	So	3. Toms, Debbie	F	So
4. Oeler, George	M	So	4. Unger, Roberta	F	So
5. Thomas, Jeff	M	So	5. Zibert, Mary	F	So

Step 3 Determine how many students need to be selected from each subgroup to have a proportional representation of each subgroup in the sample. There are four groups, and since a total of eight students is needed for the sample, two students must be selected from each subgroup.

Step 4 Select two students from each group by using random numbers. In this case, the random numbers are as follows:

Group 1	Students 5 and 4	Group 2	Students 5 and 2
Group 3	Students 1 and 3	Group 4	Students 3 and 4

The stratified sample then consists of the following people:

Peterson, John	M	Fr	Smith, Nancy	F	Fr
Meloski, Gary	M	Fr	Carson, Susan	F	Fr
Brown, Danny	M	So	Toms, Debbie	F	So
Lyons, Larry	M	So	Unger, Roberta	F	So

The major advantage of stratification is that it ensures representation of all population subgroups that are important to the study. There are two major drawbacks to stratification, however. First, if there are many variables of interest, dividing a large population into representative subgroups requires a great deal of effort. Second, if the variables are somewhat complex or ambiguous (such as beliefs, attitudes, or prejudices), it is difficult to separate individuals into the subgroups according to these variables.

Cluster Sampling

A **cluster sample** is a sample obtained by selecting a preexisting or natural group, called a *cluster,* and using the members in the cluster for the sample.

For example, many studies in education use already existing classes, such as the seventh grade in Wilson Junior High School. The voters of a certain electoral district might be surveyed to determine their preferences for a mayoral candidate in the upcoming election. Or the residents of an entire city block might be polled to ascertain the percentage of households that have two or more incomes. In cluster sampling, researchers may use all units of a cluster if that is feasible, or they may select only part of a cluster to use as a sample. This selection is done by random methods.

There are three advantages to using a cluster sample instead of other types of samples: (1) A cluster sample can reduce costs, (2) it can simplify fieldwork, and (3) it is convenient. For example, in a dental study involving X-raying fourth-grade students' teeth to see how many cavities each child had, it would be a simple matter to select a single classroom and bring the X-ray equipment to the school to conduct the study. If other sampling methods were used, researchers might have to transport the machine to several different schools or transport the pupils to the dental office.

The major disadvantage of cluster sampling is that the elements in a cluster may not have the same variations in characteristics as elements selected individually from a population. The reason is that groups of people may be more homogeneous (alike) in specific clusters such as neighborhoods or clubs. For example, the people who live in a certain neighborhood tend to have similar incomes, drive similar cars, live in similar houses, and, for the most part, have similar habits.

Speaking of
Statistics

In this study, the researchers found that subjects did better on fill-in-the-blank questions than on multiple-choice questions. Do you agree with the professor's statement, "Trusting your first impulse is your best strategy?" Explain your answer.

TESTS

Is That Your Final Answer?

Beating game shows takes more than smarts: Contestants must also overcome self-doubt and peer pressure. Two new studies suggest today's hottest game shows are particularly challenging because the very mechanisms employed to help contestants actually lead them astray.

Multiple-choice questions are one such offender, as alternative answers seem to make test-takers ignore gut instincts. To learn why, researchers at Southern Methodist University (SMU) gave two identical tests: one using multiple-choice questions and the other fill-in-the-blank. The results, recently published in the *Journal of Educational Psychology*, show that test-takers were incorrect more often when given false alternatives, and that the longer they considered those alternatives, the more credible the answers looked.

"If you sit and stew, you forget that you know the right answer," says Alan Brown, Ph.D., a psychology professor at SMU. "Trusting your first impulse is your best strategy."

Audiences can also be trouble, says Jennifer Butler, Ph.D., a Wittenberg University psychology professor. Her recent study in the *Journal of Personality and Social Psychology* found that contestants who see audience participation as peer pressure slow down to avoid making embarrassing mistakes. But this strategy backfires, as more contemplation produces more wrong answers. Worse, Butler says, if perceived peer pressure grows unbearable, contestants may opt out of answering at all, "thinking that it's better to stop than to have your once supportive audience come to believe you're an idiot."

— *Sarah Smith*

Other Types of Sampling Techniques

In addition to the four basic sampling methods, other methods are sometimes used. In **sequence sampling,** which is used in quality control, successive units taken from production lines are sampled to ensure that the products meet certain standards set by the manufacturing company.

In **double sampling,** a very large population is given a questionnaire to determine those who meet the qualifications for a study. After the questionnaires are reviewed, a second, smaller population is defined. Then a sample is selected from this group.

In **multistage sampling,** the researcher uses a combination of sampling methods. For example, suppose a research organization wants to conduct a nationwide survey for a new product being manufactured. A sample can be obtained by using the following combination of methods. First the researchers divide the 50 states into four or five regions (or clusters). Then several states from each region are selected at random. Next the states are divided into various areas by using large cities and small towns. Samples of these areas are then selected. Next, each city and town is divided into districts or wards. Finally, streets in these wards are selected at random, and the families living on these streets are given samples of the product to test and are asked to report the results. This hypothetical example illustrates a typical multistage sampling method.

The steps for conducting a sample survey are given in the Procedure Table.

Interesting Fact

Folks in extra-large aerobics classes–those with 70 to 90 participants–show up more often and are more fond of their classmates than exercisers in sessions of 18 to 26 people, report researchers at the University of Arizona.

Procedure Table

Conducting a Sample Survey

Step 1 Decide what information is needed.

Step 2 Determine how the data will be collected (phone interview, mail survey, etc.).

Step 3 Select the information-gathering instrument or design the questionnaire if one is not available.

Step 4 Set up a sampling list, if possible.

Step 5 Select the best method for obtaining the sample (random, systematic, stratified, cluster, or other).

Step 6 Conduct the survey and collect the data.

Step 7 Tabulate the data.

Step 8 Conduct the statistical analysis.

Step 9 Report the results.

Applying the Concepts 14–2

The White or Wheat Bread Debate

Read the following study and answer the questions.

A baking company selected 36 women weighing different amounts and randomly assigned them to four different groups. The four groups were white bread only, brown bread only, low-fat white bread only, and low-fat brown bread only. Each group could eat only the type of bread assigned to the group. The study lasted for eight weeks. No other changes in any of the women's diets were allowed. A trained evaluator was used to check for any differences in the women's diets. The results showed that there were no differences in weight gain between the groups over the eight-week period.

1. Did the researchers use a population or a sample for their study?
2. Based on who conducted this study, would you consider the study to be biased?
3. Which sampling method do you think was used to obtain the original 36 women for the study (random, systematic, stratified, or clustered)?
4. Which sampling method would you use? Why?
5. How would you collect a random sample for this study?
6. Does random assignment help representativeness the same as random selection does? Explain.

See page 736 for the answers.

Exercises 14–2

1. Name the four basic sampling techniques.

2. Why are samples used in statistics?

3. What is the basic requirement for a sample?

4. Why should random numbers be used when one is selecting a random sample?

5. List three incorrect methods that are often used to obtain a sample.

Figure 14–8

Student Survey at Utopia University (for Exercises 11 through 15)

Student number	Gen-der	Class rank	GPA	Miles traveled to school	IQ	Major field	Student number	Gen-der	Class rank	GPA	Miles traveled to school	IQ	Major field
1	M	Fr	1.4	1	104	Bio	26	M	Fr	1.1	8	100	Ed
2	M	Fr	2.3	2	95	Ed	27	F	Jr	2.1	3	101	Bus
3	M	So	2.7	6	108	Psy	28	M	Gr	3.7	5	99	Bio
4	F	So	3.2	7	119	Eng	29	M	Se	2.4	8	105	Eng
5	F	Gr	3.8	12	114	Ed	30	M	So	2.1	15	108	Bus
6	M	Jr	4.0	13	91	Psy	31	M	Gr	3.9	2	112	Ed
7	F	Jr	3.0	2	106	Eng	32	F	Jr	2.4	4	111	Psy
8	M	Jr	3.3	6	100	Bio	33	M	Se	2.7	6	107	Eng
9	F	Se	2.7	9	102	Eng	34	F	So	2.5	1	104	Bio
10	F	So	2.3	5	99	Ed	35	M	Se	3.2	3	96	Bus
11	M	Se	1.6	18	100	Bus	36	F	Fr	3.4	7	98	Bio
12	M	Gr	3.2	7	105	Psy	37	M	Gr	3.6	14	105	Ed
13	F	Gr	3.8	3	103	Bus	38	M	Jr	3.8	4	115	Psy
14	F	Se	3.1	5	97	Eng	39	F	Se	2.2	8	113	Eng
15	F	Jr	2.7	5	106	Bio	40	F	So	2.0	8	103	Psy
16	F	Fr	1.4	4	114	Bus	41	F	Fr	2.3	9	103	Eng
17	M	So	3.6	17	102	Ed	42	F	Se	2.5	10	99	Bus
18	M	Fr	2.2	1	101	Psy	43	M	Gr	3.7	13	114	Ed
19	F	Gr	4.0	7	108	Bus	44	M	Fr	3.0	11	121	Bus
20	M	Jr	2.1	4	97	Ed	45	M	Jr	2.1	10	101	Eng
21	F	Fr	2.0	3	113	Bio	46	F	Jr	3.4	2	104	Ed
22	F	So	3.6	4	104	Bio	47	M	So	3.6	9	105	Psy
23	F	Gr	3.3	16	110	Eng	48	M	Se	2.1	1	97	Psy
24	F	Se	2.5	4	99	Psy	49	F	Gr	3.3	12	111	Bio
25	M	So	3.0	5	96	Psy	50	F	Fr	2.2	11	102	Bio

6. What is the principle behind random numbers?

7. List the advantages and disadvantages of random sampling.

8. List the advantages and disadvantages of systematic sampling.

9. List the advantages and disadvantages of stratified sampling.

10. List the advantages and disadvantages of cluster sampling.

Using the student survey at Utopia University, shown in Figure 14–8, as the population, complete Exercises 11 through 15.

11. Using the table of random numbers in Figure 14–1, select 10 students and find the sample mean (average) of the GPA, IQ, and distance traveled to school. Compare these sample means with the population means.

12. Select a sample of 10 students by the systematic method, and compute the sample means of the GPA, IQ, and distance traveled to school of this sample. Compare these sample means with the population means.

13. Select a cluster of 10 students, for example, students 9 through 18, and compute the sample means of their GPA, IQ, and distance traveled to school. Compare these sample means with the population means.

14. Divide the 50 students into subgroups according to class rank. Then select a sample of 2 students from each rank and compute the means of these 10 students for the GPA, IQ, and distance traveled to school each day. Compare these sample means with the population means.

15. In your opinion, which sampling method(s) provided the best sample to represent the population?

Figure 14–9 shows the 50 states and the number of electoral votes each state has in the Presidential election. Using this listing as a population, complete Exercises 16 through 19.

16. Select a random sample of 10 states and find the mean number of electoral votes for this sample. Compare this mean with the population mean.

Figure 14–9

States and Number of Electoral Votes for Each (for Exercises 16 through 19)

1. Alabama	9	14. Indiana	12	27. Nebraska	5	40. South Carolina	8
2. Alaska	3	15. Iowa	8	28. Nevada	4	41. South Dakota	3
3. Arizona	7	16. Kansas	7	29. New Hampshire	4	42. Tennessee	11
4. Arkansas	6	17. Kentucky	9	30. New Jersey	16	43. Texas	29
5. California	47	18. Louisiana	10	31. New Mexico	5	44. Utah	5
6. Colorado	8	19. Maine	4	32. New York	36	45. Vermont	3
7. Connecticut	8	20. Maryland	10	33. North Carolina	13	46. Virginia	12
8. Delaware	3	21. Massachusetts	13	34. North Dakota	3	47. Washington	10
9. Florida	21	22. Michigan	20	35. Ohio	23	48. West Virginia	6
10. Georgia	12	23. Minnesota	10	36. Oklahoma	8	49. Wisconsin	11
11. Hawaii	4	24. Mississippi	7	37. Oregon	7	50. Wyoming	3
12. Idaho	4	25. Missouri	11	38. Pennsylvania	25		
13. Illinois	24	26. Montana	4	39. Rhode Island	4		

Source: *The Saturday Evening Post,* BFL&MS, Inc.

17. Select a systematic sample of 10 states and compute the mean number of electoral votes for the sample. Compare this mean with the population mean.

18. Divide the 50 states into five subgroups by geographic location, using a map of the United States. Each subgroup should include 10 states. The subgroups should be northeast, southeast, central, northwest, and southwest. Select two states from each subgroup, and find the mean number of electoral votes for the sample. Compare these means with the population mean.

19. Select a cluster of 10 states and compute the mean number of electoral votes for the sample. Compare this mean with the population mean.

20. Many research studies described in newspapers and magazines do not report the sample size or the sampling method used. Try to find a research article that gives this information; state the sampling method that was used and the sample size.

Technology *Step by Step*

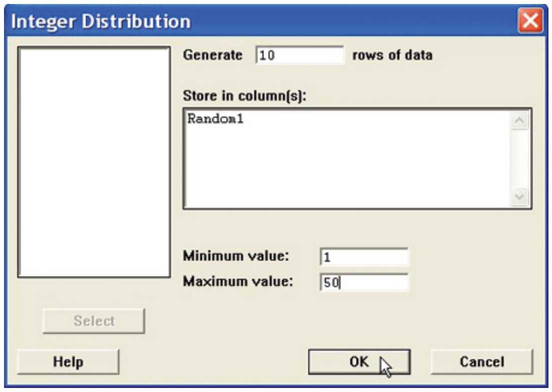
MINITAB
Step by Step

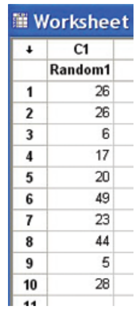
Select a Random Sample with Replacement

A simple random sample selected with replacement allows some values to be used more than once, duplicates. In the first example, a random sample of integers will be selected with replacement.

1. Select **Calc>Random Data>Integer.**
2. Type **10** for rows of data.
3. Type the name of a column, Random1, in the box for Store in column(s).
4. Type **1** for Minimum and **50** for Maximum, then click [OK].

A sample of 10 integers between 1 and 50 will be displayed in the first column of the worksheet. Every list will be different.

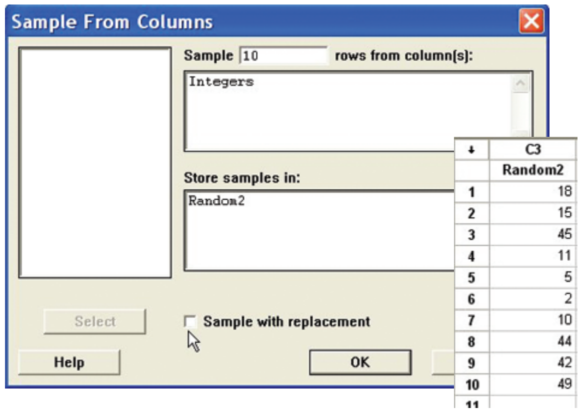




Select a Random Sample Without Replacement

To sample without replacement, make a list of integers and sample from the columns.

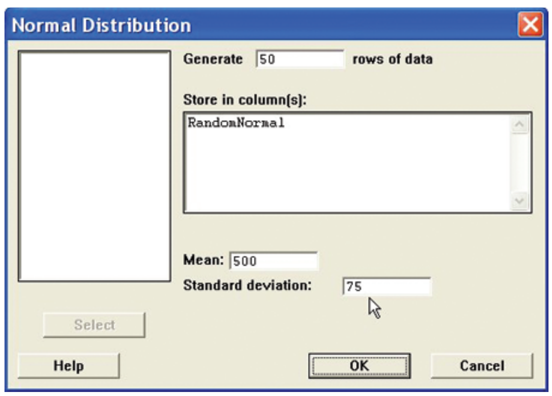
1. Select **Calc>Make Patterned Data>Simple Set of Numbers.**
2. Type **Integers** in the text box for Store patterned data in.
3. Type **1** for Minimum and **50** for Maximum. Leave 1 for steps and click [OK]. A list of the integers from 1 to 50 will be created in the worksheet.
4. Select **Calc>Random Data>Sample from columns.**
5. Sample **10** for the number of rows and Integers for the name of the column.
6. Type Random2 as the name of the new column. Be sure to leave the option for Sample with replacement unchecked.
7. Click [OK]. The new sample will be in the worksheet. There will be no duplicates.



Select a Random Sample from a Normal Distribution

No data are required in the worksheet.

1. Select **Calc>Random Data>Normal . . .**
2. Type **50** for the number of rows.
3. Press TAB or click in the box for Store in columns. Type in RandomNormal.
4. Type in **500** for the Mean and **75** for the Standard deviation.
5. Click [OK]. The random numbers are in a column of the worksheet. The distribution is sampled "with replacement." However, duplicates are not likely since this distribution is continuous. They are displayed to 3 decimal places, but many more places are stored.



Click in any cell such as row 5 of C4 RandomNormal, and you will see more decimal places.

6. To display the list, select **Data>Display data,** then select C1 RandomNormal and click [OK]. They are displayed in the same order they were selected, but going across not down.

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

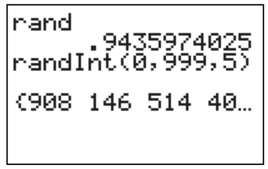
Generate Random Numbers

To generate random numbers from 0 to 1 by using the TI-83 Plus or TI-84 Plus:

1. Press **MATH** and move the cursor to PRB and press **1** for rand, then press **ENTER.** The calculator will generate a random decimal from 0 to 1.
2. To generate additional random numbers press **ENTER.**

To generate a list of random integers between two specific values:

1. Press **MATH** and move the cursor to PRB.
2. Press **5** for randInt(.
3. Enter the lowest value followed by a comma, then the largest value followed by a comma, then the number of random numbers desired followed by).
   Press **ENTER.**



Example: Generate five three-digit random numbers.

Enter **0, 999, 5)** at the randInt(as shown.

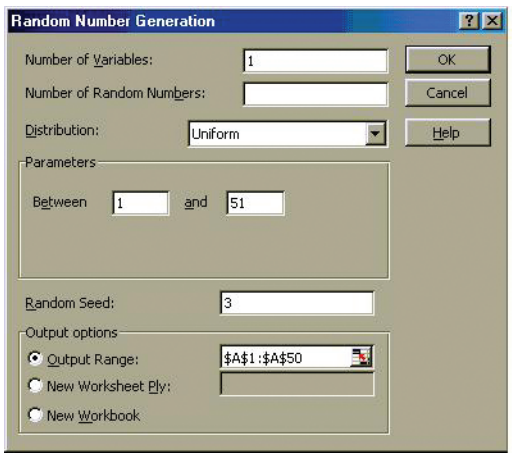
The calculator will generate five three-digit random numbers. Use the arrow keys to view the entire list.

Excel
Step by Step

Generate Random Numbers

The Data Analysis Add-In in Excel has a feature to generate random numbers from a specific probability distribution. For this example, a list of 50 random real numbers will be generated from a uniform distribution. The real numbers will then be rounded to integers between 1 and 50.

1. Open a new worksheet and select **Tool>Data Analysis>Random Number Generation** from Analysis Tools. Click [OK].
2. In the dialog box, type **1** for the Number of Variables. Leave the Number of Random Numbers box blank.
3. For Distribution, select Uniform.
4. In the Parameters box, type **1** for the lower bound and **51** for the upper bound.
5. You may type in an integer value between 1 and 51 for the Random Seed. For this example, type **3** for the Random Seed.
6. Select Output Range and type in **A1:A50.** Click [OK].



To convert the random numbers to a list of integers:

7. Select cell B1 and select the Paste Function icon from the toolbar.
8. Select the Math & Trig Function category and scroll to the Function name INT to convert the data in column A to integers.

Note that the INT function rounds the argument (input) down to the nearest integer.

9. Type cell **A1** for the Number in the INT dialog box. Click [OK].
10. While cell B1 is selected in the worksheet, move the pointer to the lower right-hand corner of the cell until a thick plus sign appears. Right-click on the mouse and drag the plus down to cell B50; then release the mouse key.
11. The numbers from column A should have been rounded to integers in column B.

Here is a sample of the data produced from the preceding procedures.

```
1.073244     1
11.98056    11
15.18195    15
14.87219    14
11.72878    11
36.97674    36
28.01193    28
36.86383    36
42.53111    42
19.56746    19
```

14–3 Surveys and Questionnaire Design

Objective 2

Recognize faulty questions on a survey and other factors that can bias responses.

Many statistical studies obtain information from *surveys.* A survey is conducted when a sample of individuals is asked to respond to questions about a particular subject. There are two types of surveys: interviewer-administered and self-administered. Interviewer-administered surveys require a person to ask the questions. The interview can be conducted face to face in an office, on a street, or in the mall, or via telephone.

Self-administered surveys can be done by mail or in a group setting such as a classroom.

When analyzing the results of surveys, you should be very careful about the interpretations. The way a question is phrased can influence the way people respond. For example, when a group of people were asked if they favored a waiting period and background check before guns could be sold, 91% of the respondents were in favor of it and 7% were against it. However, when asked if there should be a national gun registration program costing about 20% of all dollars spent on crime control, only 33% of the respondents were in favor of it and 61% were against it.

As you can see, by phrasing questions in different ways, different responses can be obtained, since the purpose of a national gun registry would include a waiting period and a background check.

When you are writing questions for a questionnaire, it is important to avoid these common mistakes.

1. *Asking biased questions.* By asking questions in a certain way, the researcher can lead the respondents to answer in the way he or she wants them to. For example, asking a question such as “Are you going to vote for the candidate Jones even though the latest survey indicates that he will lose the election?” instead of “Are you going to vote for candidate Jones?” may dissuade some people from answering in the affirmative.
2. *Using confusing words.* In this case, the participant misinterprets the meaning of the words and answers the questions in a biased way. For example, the question “Do you think people would live longer if they were on a diet?” could be misinterpreted since there are many different types of diets—weight loss diets, low-salt diets, medically prescribed diets, etc.

3. *Asking double-barreled questions.* Sometimes questions contain compound sentences that require the participant to respond to two questions at the same time. For example, the question "Are you in favor of a special tax to provide national health care for the citizens of the United States?" asks two questions: "Are you in favor of a national health care program?" and "Do you favor a tax to support it?"
4. *Using double negatives in questions.* Questions with double negatives can be confusing to the respondents. For example, the question "Do you feel that it is not appropriate to have areas where people cannot smoke?" is very confusing since *not* is used twice in the sentence.
5. *Ordering questions improperly.* By arranging the questions in a certain order, the researcher can lead the participant to respond in a way that he or she may otherwise not have done. For example, a question might ask the respondent, "At what age should an elderly person not be permitted to drive?" A later question might ask the respondent to list some problems of elderly people. The respondent may indicate that transportation is a problem based on reading the previous question.

Other factors can also bias a survey. For example, the participant may not know anything about the subject of the question but will answer the question anyway to avoid being considered uninformed. For example, many people might respond yes or no to the following question: "Would you be in favor of giving pensions to the widows of unknown soldiers?" In this case, the question makes no sense since if the soldiers were unknown, their widows would also be unknown.

Many people will make responses on the basis of what they think the person asking the questions wants to hear. For example, if a question states, "How often do you lie?" people may *understate* the incidences of their lying.

Participants will, in some cases, respond differently to questions depending on whether their identity is known. This is especially true if the questions concern sensitive issues such as income, sexuality, and abortion. Researchers try to ensure confidentiality (i.e., keeping the respondent's identity secret) rather than anonymity (soliciting unsigned responses); however, many people will be suspicious in either case.

Still other factors that could bias a survey include the time and place of the survey and whether the questions are open-ended or closed-ended. The time and place where a survey is conducted can influence the results. For example, if a survey on airline safety is conducted immediately after a major airline crash, the results may differ from those obtained in a year in which no major airline disasters occurred.

Finally, the type of questions asked influences the responses. In this case, the concern is whether the question is open-ended or closed-ended.

Unusual Stat

Of people who are struck by lightning, 85% are men.

An *open-ended question* would be one such as "List three activities that you plan to spend more time on when you retire." A *closed-ended question* would be one such as "Select three activities that you plan to spend more time on after you retire: traveling; eating out; fishing, hunting; exercising; visiting relatives."

One problem with a closed-ended question is that the respondent is forced to choose the answers that the researcher gives and cannot supply his or her own. But there is also a problem with open-ended questions in that the results may be so varied that attempting to summarize them might be difficult, if not impossible. Hence, you should be aware of what types of questions are being asked before you draw any conclusions from the survey.

There are several other things to consider when you are conducting a study that uses questionnaires. For example, a pilot study should be done to test the design and usage of the questionnaire (i.e., the *validity* of the questionnaire). The pilot study helps the researcher to pretest the questionnaire to determine if it meets the objectives of the study. It also helps the researcher to rewrite any questions that may be misleading, ambiguous, etc.

If the questions are being asked by an interviewer, some training should be given to that person. If the survey is being done by mail, a cover letter and clear directions should accompany the questionnaire.

Questionnaires help researchers to gather needed statistical information for their studies; however, much care must be given to proper questionnaire design and usage; otherwise, the results will be unreliable.

Applying the Concepts 14–3

Smoking Bans and Profits

Assume you are a restaurant owner and are concerned about the recent bans on smoking in public places. Will your business lose money if you do not allow smoking in your restaurant? You decide to research this question and find two related articles in regional newspapers. The first article states that randomly selected restaurants in Derry, Pennsylvania, that have completely banned smoking have lost 25% of their business. In that study, a survey was used and the owners were asked how much business they thought they lost. The survey was conducted by an anonymous group. It was reported in the second article that there had been a modest increase in business among restaurants that banned smoking in that same area. Sales receipts were collected and analyzed against last year's profits. The second survey was conducted by the Restaurants Business Association.

1. How has the public smoking ban affected restaurant business in Derry, Pennsylvania?
2. Why do you think the surveys reported conflicting results?
3. Should surveys based on anecdotal responses be allowed to be published?
4. Can the results of a sample be representative of a population and still offer misleading information?
5. How critical is measurement error in survey sampling?

See page 737 for the answers.

Exercises 14–3

Exercises 1 through 8 include questions that contain a flaw. Identify the flaw and rewrite the question, following the guidelines presented in this section.

1. Will you continue to shop at XYZ Department Store even though it does not carry brand names?

2. Would you buy an ABC car even if you knew the manufacturer used imported parts?

3. Should banks charge their checking account customers a fee to balance their checkbooks when customers are not able to do so?

4. Do you feel that it is not appropriate for shopping malls to have activities for children who cannot read?

5. How long have you studied for this examination?

6. Do you think children would watch less television if they read more?

7. If a plane were to crash on the border of New York and New Jersey, where should the survivors be buried?

8. Are you in favor of imposing a tax on tobacco to pay for health care related to diseases caused by smoking?

9. Find a study that uses a questionnaire. Select any questions that you feel are improperly written.

10. Many television and radio stations have a phone vote poll. If there is one in your area, select a specific day and write a brief paragraph stating the question of the day and state if it could be misleading in any way.

14–4 Simulation Techniques

Many real-life problems can be solved by employing simulation techniques.

A **simulation technique** uses a probability experiment to mimic a real-life situation.

Instead of studying the actual situation, which might be too costly, too dangerous, or too time-consuming, scientists and researchers create a similar situation but one that is less expensive, less dangerous, or less time-consuming. For example, NASA uses space shuttle flight simulators so that its astronauts can practice flying the shuttle. Most video games use the computer to simulate real-life sports such as boxing, wrestling, baseball, and hockey.

Simulation techniques go back to ancient times when the game of chess was invented to simulate warfare. Modern techniques date to the mid-1940s when two physicists, John Von Neumann and Stanislaw Ulam, developed simulation techniques to study the behavior of neutrons in the design of atomic reactors.

Mathematical simulation techniques use probability and random numbers to create conditions similar to those of real-life problems. Computers have played an important role in simulation techniques, since they can generate random numbers, perform experiments, tally the outcomes, and compute the probabilities much faster than human beings. The basic simulation technique is called the *Monte Carlo method.* This topic is discussed next.

14–5 The Monte Carlo Method

Objective 3

Solve problems, using simulation techniques.

The **Monte Carlo method** is a simulation technique using random numbers. Monte Carlo simulation techniques are used in business and industry to solve problems that are extremely difficult or involve a large number of variables. The steps for simulating real-life experiments in the Monte Carlo method are as follows:

1. List all possible outcomes of the experiment.
2. Determine the probability of each outcome.
3. Set up a correspondence between the outcomes of the experiment and the random numbers.
4. Select random numbers from a table and conduct the experiment.
5. Repeat the experiment and tally the outcomes.
6. Compute any statistics and state the conclusions.

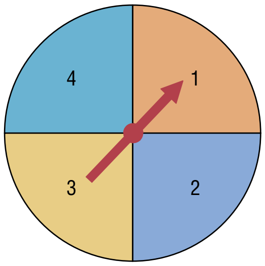
Before examples of the complete simulation technique are given, an illustration is needed for step 3 (set up a correspondence between the outcomes of the experiment and the random numbers). Tossing a coin, for instance, can be simulated by using random numbers as follows: Since there are only two outcomes, heads and tails, and since each outcome has a probability of $\frac{1}{2}$, the odd digits (1, 3, 5, 7, and 9) can be used to represent a head, and the even digits (0, 2, 4, 6, and 8) can represent a tail.

Suppose a random number 8631 is selected. This number represents four tosses of a single coin and the results T, T, H, H. Or this number could represent one toss of four coins with the same results.

An experiment of rolling a single die can also be simulated by using random numbers. In this case, the digits 1, 2, 3, 4, 5, and 6 can represent the number of spots that appear on the face of the die. The digits 7, 8, 9, and 0 are ignored, since they cannot be rolled.

Figure 14–10

Spinner with Four Numbers



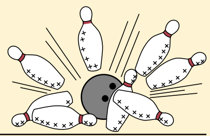
When two dice are rolled, two random digits are needed. For example, the number 26 represents a 2 on the first die and a 6 on the second die. The random number 37 represents a 3 on the first die, but the 7 cannot be used, so another digit must be selected. As another example, a three-digit daily lotto number can be simulated by using three-digit random numbers. Finally, a spinner with four numbers, as shown in Figure 14–10, can be simulated by letting the random numbers 1 and 2 represent 1 on the spinner, 3 and 4 represent 2 on the spinner, 5 and 6 represent 3 on the spinner, and 7 and 8 represent 4 on the spinner, since each number has a probability of $\frac{1}{4}$ of being selected. The random numbers 9 and 0 are ignored in this situation.

Many real-life games, such as bowling and baseball, can be simulated by using random numbers, as shown in Figure 14–11.

Figure 14–11

Example of Simulation of a Game

Source: Albert Shuylte, "Simulated Bowling Game," Student Math Notes, March 1986. Published by the National Council of Teachers of Mathematics. Reprinted with permission.



Simulated Bowling Game

Let's use the random digit table to simulate a bowling game. Our game is much simpler than commercial simulation games.

First Ball		Second Ball			
		2-Pin Split		No split	
Digit	Results	Digit	Results	Digit	Results
1–3	Strike	1	Spare	1–3	Spare
4–5	2-pin split	2–8	Leave one pin	4–6	Leave 1 pin
6–7	9 pins down	9–0	Miss both pins	7–8	*Leave 2 pins
8	8 pins down			9	+Leave 3 pins
9	7 pins down			0	Leave all pins
0	6 pins down				

*If there are fewer than 2 pins, result is a spare.
+If there are fewer than 3 pins, those pins are left.

Here's how to score bowling:

1. There are 10 frames to a **game** or **line**.
2. You roll two balls for each frame, unless you knock all the pins down with the first ball (a **strike**).
3. Your score for a frame is the sum of the pins knocked down by the two balls, if you don't knock down all 10.
4. If you knock all 10 pins down with two balls (a **spare**, shown as ◪), your score is 10 pins plus the number knocked down with the next ball.
5. If you knock all 10 pins down with the first ball (a **strike**, shown as ⊠), your score is 10 pins plus the number knocked down by the next **two** balls.
6. A **split** (shown as 0) is when there is a big space between the remaining pins. Place in the circle the number of pins remaining after the second ball.
7. A **miss** is shown as —.

Here is how one person simulated a bowling game using the random digits 7 2 7 4 8 2 2 3 6 1 6 0 4 6 1 5 5, chosen in that order from the table.

Frame

	1	2	3	4	5	6	7	8	9	10	
Digit(s)	7/2	7/4	8/2	2	3	6/1	6/0	4/6	1	5/5	
Bowling result	9 ◪ 19	9 — 28	8 ◪ 48	⊠ 77	⊠ 97	9 ◪ 116	9 — 125	8 1 134	⊠ 153	8 1 162	162

Now you try several.

Frame

	1	2	3	4	5	6	7	8	9	10	
Digit(s)											
Bowling result											

	1	2	3	4	5	6	7	8	9	10	
Digit(s)											
Bowling result											

If you wish to, you can change the probabilities in the simulation to better reflect *your* actual bowling ability.

Example 14–4 Using random numbers, simulate the gender of children born.

Solution

There are only two possibilities, female and male. Since the probability of each outcome is 0.5, the odd digits can be used to represent male births and the even digits to represent female births.

Example 14–5 Using random numbers, simulate the outcomes of a tennis game between Bill and Mike, with the additional condition that Bill is twice as good as Mike.

Solution

Unusual Stats

The average 6-year-old laughs 300 times a day; the average adult, just 17.

Since Bill is twice as good as Mike, he will win approximately two games for every one Mike wins; hence, the probability that Bill wins will be $\frac{2}{3}$, and the probability that Mike wins will be $\frac{1}{3}$. The random digits 1 through 6 can be used to represent a game Bill wins; the random digits 7, 8, and 9 can be used to represent Mike's wins. The digit 0 is disregarded. Suppose they play five games, and the random number 86314 is selected. This number means that Bill won games 2, 3, 4, and 5 and Mike won the first game. The sequence is

8	**6**	**3**	**1**	**4**
M	B	B	B	B

More complex problems can be solved by using random numbers, as shown in Examples 14–6 to 14–8.

Example 14–6 A die is rolled until a 6 appears. Using simulation, find the average number of rolls needed. Try the experiment 20 times.

Solution

Step 1 List all possible outcomes. They are 1, 2, 3, 4, 5, 6.

Step 2 Assign the probabilities. Each outcome has a probability of $\frac{1}{6}$.

Step 3 Set up a correspondence between the random numbers and the outcome. Use random numbers 1 through 6. Omit the numbers 7, 8, 9, and 0.

Step 4 Select a block of random numbers, and count each digit 1 through 6 until the first 6 is obtained. For example, the block 857236 means that it takes 4 rolls to get a 6.

8	5	7	2	3	6
	↑		↑	↑	↑
	5		2	3	6

Interesting Fact

A recent survey of more than 300 Californians ranked exercise as the surest way out of a bad mood. Listening to music was a close second.

Step 5 Repeat the experiment 19 more times and tally the data as shown.

Trial	Random number	Number of rolls
1	8 5 7 2 3 6	4
2	2 1 0 4 8 0 1 5 1 1 0 1 5 3 6	11
3	2 3 3 6	4
4	2 4 1 3 0 4 8 3 6	7
5	4 2 1 6	4
6	3 7 5 2 0 3 9 8 7 5 8 1 8 3 7 1 6	9
7	7 7 9 2 1 0 6	3
8	9 9 5 6	2
9	9 6	1
10	8 9 5 7 9 1 4 3 4 2 6	7
11	8 5 4 7 5 3 6	5
12	2 8 9 1 8 6	3
13	6	1
14	0 9 4 2 9 9 3 9 6	4
15	1 0 3 6	3
16	0 7 1 1 9 9 7 3 3 6	5
17	5 1 0 8 5 1 2 7 6	6
18	0 2 3 6	3
19	0 1 0 1 1 5 4 0 9 2 3 3 3 6	10
20	5 2 1 6	4
	Total	96

Step 6 Compute the results and draw a conclusion. In this case, one must find the average.

$$\overline{X} = \frac{\Sigma X}{n} = \frac{96}{20} = 4.8$$

Hence, the average is about 5 rolls.

Note: The theoretical average obtained from the expected value formula is 6. If this experiment is done many times, say 1000 times, the results should be closer to the theoretical results.

Example 14–7

A person selects a key at random from four keys to open a lock. Only one key fits. If the first key does not fit, she tries other keys until one fits. Find the average of the number of keys a person will have to try to open the lock. Try the experiment 25 times.

Solution

Assume that each key is numbered from 1 through 4 and that key 2 fits the lock. Naturally, the person doesn't know this, so she selects the keys at random. For the simulation, select a sequence of random digits, using only 1 through 4, until the digit 2 is reached. The trials are shown here.

Trial	Random digit (key)	Number	Trial	Random digit (key)	Number
1	2	1	14	2	1
2	2	1	15	4 2	2
3	1 2	2	16	1 3 2	3
4	1 4 3 2	4	17	1 2	2
5	3 2	2	18	2	1
6	3 1 4 2	4	19	3 4 2	3
7	4 2	2	20	2	1
8	4 3 2	3	21	2	1
9	4 2	2	22	2	1
10	2	1	23	4 2	2
11	4 2	2	24	4 3 1 2	4
12	3 1 2	3	25	3 1 2	3
13	3 1 2	3		Total	54

Next, find the average:

$$\overline{X} = \frac{\Sigma X}{n} = \frac{1 + 1 + \cdots + 3}{25} = \frac{54}{25} = 2.16$$

The theoretical average is 2.2. Again, only 25 repetitions were used; more repetitions should give a result closer to the theoretical average.

Example 14–8

A box contains five \$1 bills, three \$5 bills, and two \$10 bills. A person selects a bill at random. What is the expected value of the bill? Perform the experiment 25 times.

Solution

Step 1 List all possible outcomes. They are \$1, \$5, and \$10.

Step 2 Assign the probabilities to each outcome:

$$P(\$1) = \tfrac{5}{10} \qquad P(\$5) = \tfrac{3}{10} \qquad P(\$10) = \tfrac{2}{10}$$

Step 3 Set up a correspondence between the random numbers and the outcomes. Use random numbers 1 through 5 to represent a \$1 bill being selected, 6 through 8 to represent a \$5 bill being selected, and 9 and 0 to represent a \$10 bill being selected.

Steps 4 and 5 Select 25 random numbers and tally the results.

Number	Results (\$)
4 5 8 2 9	1, 1, 5, 1, 10
2 5 6 4 6	1, 1, 5, 1, 5
9 1 8 0 3	10, 1, 5, 10, 1
8 4 0 6 0	5, 1, 10, 5, 10
9 6 9 4 3	10, 5, 10, 1, 1

Step 6 Compute the average:

$$\overline{X} = \frac{\Sigma X}{n} = \frac{\$1 + \$1 + \$5 + \cdots + \$1}{25} = \frac{\$116}{25} = \$4.64$$

Hence, the average (expected value) is \$4.64.

Recall that using the expected value formula $E(X) = \Sigma X \cdot P(X)$ gives a theoretical average of

$$E(X) = \Sigma[X \cdot P(X)] = (0.5)(\$1) + (0.3)(\$5) + (0.2)(\$10) = \$4.00$$

Remember that simulation techniques do not give exact results. The more times the experiment is performed, though, the closer the actual results should be to the theoretical results. (Recall the law of large numbers.)

The steps for solving problems using the Monte Carlo method are summarized in the Procedure Table.

Procedure Table

Simulating Experiments Using the Monte Carlo Method

Step 1 List all possible outcomes of the experiment.

Step 2 Determine the probability of each outcome.

Step 3 Set up a correspondence between the outcomes of the experiment and the random numbers.

Step 4 Select random numbers from a table and conduct the experiment.

Step 5 Repeat the experiment and tally the outcomes.

Step 6 Compute any statistics and state the conclusions.

Applying the Concepts 14–4

Simulations

Answer the following questions:

1. Define simulation technique.
2. Have simulation techniques been used for very many years?
3. Is it cost-effective to do simulation testing on some things such as airplanes or automobiles?
4. Why might simulation testing be better than real-life testing? Give examples.
5. When did physicists develop computer simulation techniques to study neutrons?
6. When could simulations be misleading or harmful? Give examples.
7. Could simulations have prevented previous disasters such as the Hindenburg or the Space Shuttle disaster?
8. What discipline is simulation theory based in?

See page 737 for the answers.

Exercises 14–5

1. Define simulation techniques.

2. Give three examples of simulation techniques.

3. Who is responsible for the development of modern simulation techniques?

4. What role does the computer play in simulation?

5. What are the steps in the simulation of an experiment?

6. What purpose do random numbers play in simulation?

7. What happens when the number of repetitions is increased?

For Exercises 8 through 13, explain how each experiment can be simulated by using random numbers.

8. A spinner contains six equal areas.

9. A basketball player makes 70% of her shots.

10. A certain brand of DVD player manufactured has a 10% defective rate.

11. An archer hits a target 80% of the time.

12. Two players match pennies.

13. Three players play odd man out. (Three coins are tossed; if all three match, the game is repeated and no one wins. If two players match, the third person wins all three coins.)

For Exercises 14 through 21, use random numbers to simulate the experiments. The number in parentheses is the number of times the experiment should be repeated.

14. A coin is tossed until four heads are obtained. Find the average number of tosses necessary. (50)

15. A die is rolled until all faces appear at least once. Find the average number of tosses. (30)

16. A caramel corn company gives four different prizes, one in each box. They are placed in the boxes at random. Find the average number of boxes a person needs to buy to get all four prizes. (40)

17. Two teams are evenly matched. They play a tournament in which the first team to win three games wins the tournament. Find the average number of games the tournament will last. (20)

18. To win a certain lotto, a person must spell the word *big.* Sixty percent of the tickets contain the letter *b,* 30% contain the letter *i,* and 10% contain the letter *g*. Find the average number of tickets a person must buy to win the prize. (30)

19. Two shooters shoot clay pigeons. Gail has an 80% accuracy rate and Paul has a 60% accuracy rate. Paul shoots first. The first person who hits the target wins. Find the probability that each wins. (30).

20. In Exercise 19, find the average number of shots fired. (30)

21. A basketball player has a 60% success rate for shooting foul shots. If she gets two shots, find the probability that she will make one or both shots. (50).

22. Select a game such as baseball or football and write a simulation using random numbers.

23. Explain how cards can be used to generate random numbers.

24. Explain how a pair of dice can be used to generate random numbers.

14–6 Summary

To obtain information and make inferences about a large population, researchers select a sample. A sample is a subgroup of the population. Using a sample rather than a population, researchers can save time and money, get more detailed information, and get information that otherwise would be impossible to obtain.

The four most common methods researchers use to obtain samples are random, systematic, stratified, and cluster sampling methods. In random sampling, some type of random method (usually random numbers) is used to obtain the sample. In systematic sampling, the researcher selects every *k*th person or item after selecting the first one at random. In stratified sampling, the population is divided into subgroups according to various characteristics, and elements are then selected at random from the subgroups. In cluster sampling, the researcher selects an intact group to use as a sample. When the population is large, multistage sampling (a combination of methods) is used to obtain a subgroup of the population.

Researchers must use caution when conducting surveys and designing questionnaires; otherwise, conclusions obtained from these will be inaccurate. Guidelines were presented in Section 14–3.

Most sampling methods use random numbers, which can also be used to simulate many real-life problems or situations. The basic method of simulation is known as the Monte Carlo method. The purpose of simulation is to duplicate situations that are too dangerous, too costly, or too time-consuming to study in real life. Most simulation techniques can be done on the computer or calculator, since they can rapidly generate random numbers, count the outcomes, and perform the necessary computations.

Sampling and simulation are two techniques that enable researchers to gain information that might otherwise be unobtainable.

Important Terms

Review Exercises

Use Figure 14–12 for Exercises 1 through 8.

1. Select a random sample of 10 people, and find the mean of the weights of the individuals. Compare this mean with the population mean.

2. Select a systematic sample of 10 people, and compute the mean of their weights. Compare this mean with the population mean.

3. Divide the individuals into subgroups of males and females. Select five individuals from each group, and find the mean of their weights. Compare these means with the population mean.

4. Select a cluster of 10 people, and find the mean of their weights. Compare this mean with the population mean.

Figure 14–12

Population for Exercises 1 through 8

Individual	Gender	Weight	Systolic blood pressure	Individual	Gender	Weight	Systolic blood pressure	Individual	Gender	Weight	Systolic blood pressure
1	F	122	132	18	F	118	125	35	M	172	116
2	F	128	116	19	F	107	138	36	M	175	123
3	M	183	140	20	M	214	121	37	F	101	114
4	M	165	136	21	F	114	127	38	F	123	113
5	M	192	120	22	M	119	125	39	M	186	145
6	F	116	118	23	F	125	114	40	F	100	119
7	M	206	116	24	M	182	137	41	M	202	135
8	F	131	120	25	F	127	127	42	F	117	121
9	M	155	118	26	F	132	130	43	F	120	130
10	F	106	122	27	M	198	114	44	M	193	125
11	F	103	119	28	F	135	119	45	M	200	115
12	M	169	136	29	M	183	137	46	F	118	132
13	M	173	134	30	F	140	123	47	F	121	143
14	M	195	145	31	M	189	135	48	M	189	128
15	F	107	113	32	M	165	121	49	M	114	118
16	M	201	111	33	M	211	117	50	M	174	138
17	F	114	141	34	F	111	127				

5. Repeat Exercise 1 for blood pressure.
6. Repeat Exercise 2 for blood pressure.
7. Repeat Exercise 3 for blood pressure.
8. Repeat Exercise 4 for blood pressure.

For Exercises 9 through 13, explain how to simulate each experiment by using random numbers.

9. A baseball player strikes out 40% of the time.
10. An airline overbooks 15% of the time.
11. Two players roll a die. The higher number wins.
12. Player 1 rolls two dice. Player 2 rolls one die. If the number on the single die matches one number of the player who rolled the two dice, player 2 wins. Otherwise, player 1 wins.
13. Two players play rock, paper, scissors. The rules are as follows: Since paper covers rock, paper wins. Since rock breaks scissors, rock wins. Since scissors cut paper, scissors win. Each person selects rock, paper, or scissors by random numbers and then compares results.

For Exercises 14 through 18, use random numbers to simulate the experiments. The number in parentheses is the number of times the experiment should be repeated.

14. A football is placed on the 10-yard line, and a team has four downs to score a touchdown. The team can move the ball only 0 to 5 yards per play. Find the average number of times the team will score a touchdown. (30)
15. In Exercise 14, find the average number of plays it will take to score a touchdown. Ignore the four-downs rule and keep playing until a touchdown is scored. (30)
16. Four dice are rolled 50 times. Find the average of the sum of the number of spots that will appear. (50)
17. A field goal kicker is successful in 60% of his kicks inside the 35-yard line. Find the probability of kicking three field goals in a row. (50)
18. A sales representative finds that there is a 30% probability of making a sale by visiting the potential customer personally. For every 20 calls, find the probability of making three sales in a row. (50)

For Exercises 19 through 22, explain what is wrong with each question. Rewrite each one following the guidelines in this chapter.

19. How often do you run red lights?
20. Do you think students who are not failing should not be tutored?
21. Do you think all automobiles should have heavy-duty bumpers, even though it will raise the price of the cars by $500?
22. Explain the difference between an open-ended question and a closed-ended question.

Data Analysis

The Data Bank is found in Appendix D.

1. From the Data Bank, choose a variable. Select a random sample of 20 individuals, and find the mean of the data.
2. Select a systematic sample of 20 individuals, and using the same variable as in Exercise 1, find the mean.
3. Select a cluster sample of 20 individuals, and using the same variable as in Exercise 1, find the mean.
4. Stratify the data according to marital status and gender, and sample 20 individuals. Compute the mean of the sample variable selected in Exercise 1 (use four groups of five individuals).
5. Compare all four means and decide which one is most appropriate. (*Hint:* Find the population mean.)

Chapter Quiz

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

1. When researchers are sampling from large populations, such as adult citizens living in the United States, they may use a combination of sampling techniques to ensure representativeness.
2. Simulation techniques using random numbers are a substitute for performing the actual statistical experiment.
3. When researchers perform simulation experiments, they do not need to use random numbers since they can make up random numbers.
4. Random samples are said to be unbiased.

Statistics Today

The Monty Hall Problem—Revisited

It appears that it does not matter whether the contestant switches doors because he is given a choice of two doors, and the chance of winning the prize is 1 out of 2, or $\frac{1}{2}$. This reasoning, however, is incorrect. Consider the three possibilities for the prize. It could be behind door A, B, or C. Also consider the fact that the contestant has selected door A. Now the three situations look like this:

	Door		
Case	**A**	**B**	**C**
1	Prize	Empty	Empty
2	Empty	Prize	Empty
3	Empty	Empty	Prize

In case 1, the contestant selected door A, and if the contestant switched after being shown that there was no prize behind either door B or door C, he'd lose. In case 2, the contestant selected door A, and Monty will open door C, so if the contestant would switch, he would win the prize. In case 3, the contestant selected door A, and Monty will open door B, so if the contestant would switch, he would win the prize. Hence, by switching, the probability of winning is $\frac{2}{3}$ and the probability of losing is $\frac{1}{3}$. The same reasoning can be used no matter which door you select.

You can simulate this problem by using three cards, say, an ace (prize) and two other cards. Have a person arrange the cards in a row and let you select a card. After the person turns over one of the cards (a nonace), then switch. Keep track of the number of times you win. You can also play this game on the Internet by going to the website http://www.stat.sc.edu/~west/javahtml/LetsMakeaDeal.html

Select the best answer.

5. When all subjects under study are used, the group is called a ________.

a. Population *c.* Sample
b. Large group *d.* Study group

6. When a population is divided into subgroups with similar characteristics and then a sample is obtained, this method is called ________ sampling.

a. Random *c.* Stratified
b. Systematic *d.* Cluster

7. Interviewing selected people at a local supermarket can be considered an example of ________ sampling.

a. Random *c.* Convenience
b. Systematic *d.* Stratified

Complete the following statements with the best answer.

8. In general, when one conducts sampling, the ________ the sample, the more representative it will be.

9. When samples are not representative, they are said to be ________.

10. When all residents of a street are interviewed for a survey, the sampling method used is ________.

Use Figure 14–12 in the Review Exercises (page 733) for Exercises 11 through 14.

11. Select a random sample of 12 people, and find the mean of the blood pressures of the individuals. Compare this with the population mean.

12. Select a systematic sample of 12 people, and compute the mean of their blood pressures. Compare this with the population mean.

13. Divide the individuals into subgroups of six males and six females. Find the means of their blood pressures. Compare these means with the population mean.

14. Select a cluster of 12 people, and find the mean of their blood pressures. Compare this with the population mean.

For Exercises 15 through 19, explain how each could be simulated by using random numbers.

15. A chess player wins 45% of his games.

16. A travel agency has a 5% cancellation rate.

17. Two players select a card from a deck with no face cards. The player who gets the higher card wins.

18. One player rolls two dice. The other player selects a card from a deck. Face cards count as 11 for a jack, 12 for a queen, and 13 for a king. The player with the higher total points wins.
19. Two players toss two coins. If they match, player 1 wins; otherwise, player 2 wins.

For Exercises 20 through 24, use random numbers to simulate the experiments. The number in parentheses is the number of times the experiment should be done.

20. A telephone solicitor finds that there is a 15% probability of selling her product over the phone. For every 20 calls, find the probability of making two sales in a row. (100)
21. A field goal kicker is successful in 65% of his kicks inside the 40-yard line. Find the probability of his kicking four field goals in a row. (40)
22. Two coins are tossed. Find the average number of times two tails will appear. (40)
23. A single card is drawn from a deck. Find the average number of times it takes to draw an ace. (30)
24. A bowler finds that there is a 30% probability that he will make a strike. For every 15 frames he bowls, find the probability of making two strikes. (30)

Critical Thinking Challenges

1. Explain why two different opinion polls might yield different results on a survey. Also, give an example of an opinion poll and explain how the data may have been collected.
2. Use a computer to generate random numbers to simulate the following real-life problem.

   In a certain geographic region, 40% of the people have type O blood. On a certain day, the blood center needs 4 pints of type O blood. On average, how many donors are needed to obtain 4 pints of type O blood?

Data Projects

Where appropriate, use MINITAB, the TI-83 Plus, the TI-84 Plus, or a computer program of your choice to complete the following exercises.

1. Using the rules given in Figure 14–11 on page 727 of your textbook, play the simulated bowling game at least 10 times. Each game consists of 10 frames.
   a. Analyze the results of the scores by finding the mean, median, mode, range, variance, and standard deviation.
   b. Draw a box plot and explain the nature of the distribution.
   c. Write several paragraphs explaining the results.
   d. Compare this simulation with real bowling. Do you think the game actually simulates bowling? Why or why not?
2. Select a sports game that you like to play or watch on television (e.g., baseball, golf, or hockey). Write a simulated version of the game, using random numbers or dice. Play the game several times and answer these questions.
   a. Does your simulated game represent the real game accurately?
   b. Is your game one of pure chance, or is strategy involved?
   c. What are some shortcomings of your game?
   d. What parts of the real game cannot be simulated in your game?
   e. Is there any way that you could improve your simulated game by changing some rules?

Answers to Applying the Concepts

Section 14–2 The White or Wheat Bread Debate

1. The researchers used a sample for their study.
2. Answers will vary. One possible answer is that we might have doubts about the validity of the study, since the baking company that conducted the experiment has an interest in the outcome of the experiment.
3. The sample was probably a convenience sample.
4. Answers will vary. One possible answer would be to use a simple random sample.

5. Answers will vary. One possible answer is that a list of women's names could be obtained from the city in which the women live. Then a simple random sample could be selected from this list.
6. The random assignment helps to spread variation among the groups. The random selection helps to generalize from the sample back to the population. These are two different issues.

Section 14–3 Smoking Bans and Profits

1. It is uncertain how public smoking bans affected restaurant business in Derry, Pennsylvania, since the survey results were conflicting.
2. Since the data were collected in different ways, the survey results were bound to have different answers. Perceptions of the owners will definitely be different from an analysis of actual sales receipts, particularly if the owners assumed that the public smoking bans would hurt business.
3. Answers will vary. One possible answer is that it would be difficult to not allow surveys based on anecdotal responses to be published. At the same time, it would be good for those publishing such survey results to comment on the limitations of these surveys.
4. We can get results from a representative sample that offer misleading information about the population.
5. Answers will vary. One possible answer is that measurement error is important in survey sampling in order to give ranges for the population parameters that are being investigated.

Section 14–4 Simulations

1. A simulation uses a probability experiment to mimic a real-life situation.
2. Simulation techniques date back to ancient times.
3. It is definitely cost-effective to run simulations for expensive items such as airplanes and automobiles.
4. Simulation testing is safer, faster, and less expensive than many real-life testing situations.
5. Computer simulation techniques were developed in the mid-1940s.
6. Answers will vary. One possible answer is that some simulations are far less harmful than conducting an actual study on the real-life situation of interest.
7. Answers will vary. Simulations could have possibly prevented disasters such as the Hindenburg or the Space Shuttle disaster. For example, data analysis after the Space Shuttle disaster showed that there was a decent chance that something would go wrong on that flight.
8. Simulation theory is based in probability theory.