and chooses the numbers 1, 2, 3, 4, 5, 6. Her friend George says that this isn’t a good choice since it is very unlikely that a random sample will turn up the first six numbers. Is he right?

Solution
No. It is true that the combination 1, 2, 3, 4, 5, 6 is unlikely, but every other combination is equally unlikely. In a simple random sample of size 6, every collection of six numbers is equally likely (or equally unlikely) to come up. So Sally has the same chance as anyone to win the jackpot.

EXAMPLE 1.5
Using technology to draw a simple random sample

Use technology to draw a simple random sample of five employees from the following list:

5. Paul Campbell 15. James Graves 25. Frank Lipka 35. Sherry Thomas

Solution
We will use the TI-84 Plus graphing calculator. The step-by-step procedure is presented in the Using Technology section on page 9. We begin by choosing a seed, which is a number that the calculator uses to get the random number generator started. Display (a) shows the seed being set to 21. (The seed can be chosen in almost any way; this number was chosen by looking at the seconds display on a digital watch.) Display (b) presents the five numbers in the sample.

\[
\begin{align*}
21 \rightarrow \text{rand} & \quad \text{randInt}\{1,40,5\} \\
\text{(a)} & \quad \text{(b)} \\
27 & \quad 39 \\
30 & \quad 35 \\
17 & \quad 17
\end{align*}
\]

The simple random sample consists of the employees with numbers 27, 39, 30, 35, and 17. These are Laverne Mitchell, Lizbet Valdez, Gary Sanders, Sherry Thomas, and Andrew Huang.

CAUTION
If you use a different type of calculator, a different statistical package, or a different seed, you will get a different random sample. This is perfectly all right. So long as the sample is drawn by using a correct procedure, it is a valid random sample.

Check Your Understanding
1. A pollster wants to estimate the proportion of voters in a certain town who are Democrats. He goes to a large shopping mall and approaches people to ask whether they are Democrats. Is this a simple random sample? Explain.

2. A telephone company wants to estimate the proportion of customers who are satisfied with their service. They use a computer to generate a list of random phone numbers and call those people to ask them whether they are satisfied. Is this a simple random sample? Explain.

Answers are on page 1.
Solution
We start with the third car, then count by fives to determine which cars will be sampled. The sample will consist of cars numbered 3, 8, 13, 18, and so on.

Voluntary response sampling
Voluntary response samples are often used by the media to try to engage the audience. For example, a news commentator will invite people to tweet an opinion, or a radio announcer will invite people to call the station to say what they think. How reliable are voluntary response samples? To put it simply, voluntary response samples are never reliable. People who go to the trouble to volunteer an opinion tend to have stronger opinions than is typical of the population. In addition, people with negative opinions are often more likely to volunteer their responses than those with positive opinions.

Figures 1.1–1.4 illustrate several valid methods of sampling.

![Figure 1.1 Simple random sampling](image1)

![Figure 1.2 Systematic sampling](image2)

![Figure 1.3 Stratified sampling](image3)

![Figure 1.4 Cluster sampling](image4)

Check Your Understanding

3. A radio talk-show host invites listeners to send an email to express their opinions on an upcoming election. More than 10,000 emails are received. What kind of sample is this? Voluntary response

4. Every 10 years, the U.S. Census Bureau attempts to count every person living in the United States. To check the accuracy of their count in a certain city, they draw a sample of census districts (roughly equivalent to a city block) and recount everyone in the sampled districts. What kind of sample is formed by the people who are recounted? Cluster
EXCEL
Drawing a simple random sample

Step 1. In Column A, enter the values 1 through the population size \( N \). For Example 1.5, \( N = 40 \).

Step 2. In Column B, next to each value in Column A, enter the command \( =\text{rand()} \). This results in a randomly generated number between 0 and 1 in each cell in Column B.

Step 3. Select all values in Columns A and B and then click on the Data menu and select Sort.

Step 4. In the Sort by field, enter Column B and select Smallest to Largest in the Order field. Press OK. Column A now contains the random sample. Our random sample begins with 17, 12, 28, 20, 6, ... (Figure E).

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.919738</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>0.83333</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>0.105403</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0.863496</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.820379</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>0.514735</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>0.951148</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>0.756971</td>
</tr>
<tr>
<td>9</td>
<td>39</td>
<td>0.826555</td>
</tr>
<tr>
<td>10</td>
<td>26</td>
<td>0.315564</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>0.119579</td>
</tr>
<tr>
<td>12</td>
<td>30</td>
<td>0.329793</td>
</tr>
</tbody>
</table>

Figure E

SECTION 1.1 Exercises

Exercises 1–6 are the Check Your Understanding exercises located within the section.

Understanding the Concepts
In Exercises 7–12, fill in each blank with the appropriate word or phrase.

7. The entire collection of individuals about which information is sought is called a ________ population.

8. A ________ is a subset of a population. sample

9. A ________ is a type of sample that is analogous to a lottery. simple random sample

10. A sample that is not drawn by a well-defined random method is called a ________. sample of convenience

11. A ________ sample is one in which the population is divided into groups and a random sample of groups is drawn. cluster

12. A ________ sample is one in which the population is divided into groups and a random sample is drawn from each group. stratified

In Exercises 13–16, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

13. A sample of convenience is never acceptable. False

14. In a cluster sample, the population is divided into groups, and a random sample from each group is drawn. False

15. Both stratified sampling and cluster sampling divide the population into groups. True

16. One reason that voluntary response sampling is unreliable is that people with stronger views tend to express them more readily. True

Practicing the Skills
In Exercises 17–20, determine whether the number described is a statistic or a parameter.

17. In a recent poll, 57% of the respondents supported a school bond issue. Statistic

18. The average age of the employees in a certain company is 35 years. Parameter

19. Of the students enrolled in a certain college, 80% are full-time. Parameter

20. In a survey of 300 high-school students, 60% of them said that they intended to go to college. Statistic

Exercises 21–24 refer to the population of animals in the following table. The population is divided into four groups: mammals, birds, reptiles, and fish.

<table>
<thead>
<tr>
<th></th>
<th>Mammals</th>
<th>Birds</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aardvark</td>
<td>Flamingo</td>
</tr>
<tr>
<td>2</td>
<td>Buffalo</td>
<td>Swan</td>
</tr>
<tr>
<td>3</td>
<td>Elephant</td>
<td>Sparrow</td>
</tr>
<tr>
<td>4</td>
<td>Squirrel</td>
<td>Parrot</td>
</tr>
<tr>
<td>5</td>
<td>Rabbit</td>
<td>Pelican</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Reptiles</th>
<th>Fish</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Geckio</td>
<td>Catfish</td>
</tr>
<tr>
<td>22</td>
<td>Iguana</td>
<td>Tuna</td>
</tr>
<tr>
<td>23</td>
<td>Chameleon</td>
<td>Cod</td>
</tr>
<tr>
<td>24</td>
<td>Rattlesnake</td>
<td>Salmon</td>
</tr>
<tr>
<td>25</td>
<td>Boa</td>
<td>Crocodile</td>
</tr>
</tbody>
</table>

21. Simple random sample: Draw a simple random sample of eight animals from the list of 40 animals in the table.
22. Another sample: Draw a sample of eight animals by drawing a simple random sample of two animals from each group. What kind of sample is this? 

23. Another sample: Draw a simple random sample of two groups of animals from the four groups, and construct a sample of 20 animals by including all the animals in the sampled groups. What kind of sample is this?

24. Another sample: Choose a random number between 1 and 5. Include the animal with that number in your sample, along with every fifth animal thereafter, to construct a sample of eight animals. What kind of sample is this?

In Exercises 25–36, identify the kind of sample that is described.

25. Parking on campus: A college faculty consists of 400 men and 250 women. The college administration wants to draw a sample of 65 faculty members to ask their opinion about a new parking fee. They draw a simple random sample of 40 men and another simple random sample of 25 women.

26. Cruising the mall: A pollster walks around a busy shopping mall, and approaches people passing by to ask them how often they shop at the mall.

27. What’s on TV? A pollster obtains a list of all the residential addresses in a certain town, and uses a computer random number generator to choose 150 of them. The pollster visits each of the 150 households and interviews all the adults in each household about their television viewing habits.

28. Don’t drink and drive: Police at a sobriety checkpoint pull over every fifth car to determine whether the driver is sober.

29. Tell us your opinion: A television newscaster invites viewers to tweet their opinions on a proposed bill on immigration policy. More than 50,000 people express their opinions in this way.

30. Reading program: The superintendent of a large school district wants to test the effectiveness of a new program designed to improve reading skills among elementary school children. There are 30 elementary schools in the district. The superintendent chooses a simple random sample of five schools, and institutes the new reading program in those schools. A total of 4700 children attend these five schools.

31. Customer survey: All the customers who entered a store on a particular day were given a survey to fill out concerning their opinions of the service at the store.

32. Raffle: Five hundred people attend a charity event, and each buys a raffle ticket. The 500 ticket stubs are put in a drum and thoroughly mixed, and 10 of them are drawn. The 10 people whose tickets are drawn win a prize.

33. Hospital survey: The director of a hospital pharmacy chooses at random 100 people age 60 or older from each of three surrounding counties to ask their opinions of a new prescription drug program.

34. Bus schedule: Officials at a metropolitan transit authority want to get input from people who use a certain bus route about a possible change in the schedule. They randomly select 5 buses during a certain week and poll all riders on those buses about the change.

35. How Much did you spend? A retailer samples 25 receipts from the past week by numbering all the receipts, generating 25 random numbers, and sampling the receipts that correspond to these numbers.

36. Phone features: A cell phone company wants to draw a sample of 600 customers to gather opinions about potential new features on upcoming phone models. The company draws a random sample of 200 from customers with iPhones, a random sample of 100 from customers with LG phones, a random sample of 100 from customers with Samsung phones, and a random sample of 200 from customers with other phones.

37. Computer network: Every third day, a computer network administrator analyzes the company’s network logs to check for signs of computer viruses.

38. Smartphone apps: A smartphone app produces a message requesting customers to click on a link to rate the app.

Working with the Concepts

39. You’re giving me a headache: A pharmaceutical company wants to test a new drug that is designed to provide superior relief from headaches. They want to select a sample of headache sufferers to try the drug. Do you think that it is feasible to draw a simple random sample of headache sufferers, or will it be necessary to use a sample of convenience? Explain your reasoning.

40. Pay more for recreation? The director of the recreation center at a large university wants to sample 100 students to ask them whether they would support an increase in their recreation fees in order to expand the hours that the center is open. Do you think that it is feasible to draw a simple random sample of students, or will it be necessary to use a sample of convenience? Explain your reasoning.

41. Voter preferences: A pollster wants to sample 500 voters in a town to ask them who they plan to vote for in an upcoming election. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.

42. Quality control: Products come off an assembly line at the rate of several hundred per hour. It is desired to sample 10% of them to check whether they meet quality standards. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.
43. **On-site day care**: A large company wants to sample 200 employees to ask their opinions about providing a day care center for the employees' children. They want to be sure to sample equal numbers of men and women. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.

44. **The tax man cometh**: The Internal Revenue Service wants to sample 1000 tax returns that were submitted last year to determine the percentage of returns that had a refund. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.

### Extending the Concepts

45. **Draw a sample**: Imagine that you are asked to determine students' opinions at your school about a potential change in library hours. Describe how you could go about getting a sample of each of the following types: simple random sample, sample of convenience, voluntary response sample, stratified sample, cluster sample, systematic sample.

46. **A systematic sample is a cluster sample**: Explain how a systematic sample is actually a type of cluster sample.

---

**Answers to Check Your Understanding Exercises for Section 1.1**

1. No; this sample consists only of people in the town who visit the mall.
2. Yes; every group of \(n\) customers, where \(n\) is the sample size, is equally likely to be chosen.
3. Voluntary response sample
4. Cluster sample
5. Stratified sample
6. Systemic sample

---

**SECTION 1.2 Types of Data**

**Objectives**

1. Understand the structure of a typical data set
2. Distinguish between qualitative and quantitative variables
3. Distinguish between ordinal and nominal variables
4. Distinguish between discrete and continuous variables

**Data Sets**

In Section 1.1, we described various methods of collecting information by sampling. Once the information has been collected, the collection is called a **data set**. A simple example of a data set is presented in Table 1.1, which shows the major, final exam score, and grade for several students in a certain statistics class.

<table>
<thead>
<tr>
<th>Table 1.1 Major, Final Exam Score, and Grade for Several Students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

Table 1.1 illustrates some basic features that are found in most data sets. Informatic is collected on **individuals**. In this example, the individuals are students. In many cases, individuals are people; in other cases, they can be animals, plants, or things. The characteristics of the individuals about which we collect information are called **variable**. In this example, the variables are major, exam score, and grade. Finally, the values of the variables that we obtain are the **data**. So, for example, the data for individual #1 are Major = Psychology, Exam score = 92, and Grade = A.
Solution

a. Age at a person’s last birthday is discrete. The possible values are 0, 1, 2, and so forth.
b. Height is continuous. A person’s height is not restricted to any list of values.
c. Number of siblings is discrete. The possible values are 0, 1, 2, and so forth.
d. Distance commuted to work is continuous. It is not restricted to any list of values.

Check Your Understanding

2. Which are qualitative and which are quantitative?
   a. The number of patients admitted to a hospital on a given day
   b. The model of car last sold by a particular car dealer
   c. The name of your favorite song
   d. The seating capacity of an auditorium

3. Which are nominal and which are ordinal?
   a. The names of the streets in a town
   b. The movie ratings G, PG, PG-13, R, and NC-17
   c. The winners of the gold, silver, and bronze medals in an Olympic swimming competition

4. Which are discrete and which are continuous?
   a. The number of female members of the U.S. House of Representatives
   b. The amount of water used by a household during a given month
   c. The number of stories in an apartment building
   d. A person’s body temperature

Answers are on page 18.

SECTION 1.2 Exercises

Exercises 1–4 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 5–10, fill in each blank with the appropriate word or phrase.

5. The characteristics of individuals about which we collect information are called _______.
6. Variables that classify individuals into categories are called _______.
7. _______ variables are always numerical.
8. Qualitative variables can be divided into two types: _______ and _______.
9. A _______ variable is a quantitative variable whose possible values can be listed.
10. _______ variables can take on any value in some interval.

In Exercises 11–14, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

11. Qualitative variables describe how much or how many of something there is.  False
12. A nominal variable is a qualitative variable with no natural ordering.  True

13. A discrete variable is one whose possible values can be listed.  True
14. A person’s height is an example of a continuous variable.  True

Practicing the Skills

In Exercises 15–24, determine whether the data described are qualitative or quantitative.

15. Your best friend’s name
16. Your best friend’s age
17. The number of touchdowns in a football game
18. The title of your statistics book
19. The number of files on a computer
20. The waist size of a pair of jeans
21. The ingredients in a recipe
22. Your school colors
23. The makes of cars sold by a particular car dealer
24. The number of cars sold by a car dealer last month

In Exercises 25–32, determine whether the data described are nominal or ordinal.

25. The categories Strongly disagree, Disagree, Neutral, Agree, and Strongly agree on a survey
26. The names of the counties in a state 
   Nominal
27. The shirt sizes of Small, Medium, Large, and X-Large 
   Ordinal
28. I got an A in statistics, a B in biology, and C's in history and English. 
   Ordinal
29. This semester, I am taking statistics, biology, history, and English. 
   Nominal
30. I ordered a pizza with pepperoni, mushrooms, olives, and onions. 
   Nominal
31. In the track meet, I competed in the high jump and the pole vault. 
   Ordinal
32. I finished first in the high jump and third in the pole vault. 
   Ordinal

In Exercises 33–40, determine whether the data described are discrete or continuous.

33. The amount of caffeine in a cup of Starbucks coffee 
   Continuous

Working with the Concepts

41. Ring tones: Following are the ten top-selling ring tones for the year 2012:
   1. Alicia Keys — Girl on Fire
   2. PSY — Gangnam Style
   3. Florida Georgia Line — Cruise
   4. Taylor Swift — I Knew You Were Trouble
   5. 2 Chainz — I'm Different
   6. Rihanna — Diamonds
   7. Gary Allan — Every Storm
   8. Little Big Town — Pontoon
   9. The Band Perry — Better Dig Two
   10. Lil Wayne — No Worries
   Source: www.billboard.com

Are these data nominal or ordinal?

42. More Ring tones: The following table presents the number of weeks that each of the ring tones in Exercise 41 spent on the top-ten chart as of February 2012.

<table>
<thead>
<tr>
<th>Ringtone</th>
<th>Weeks in top ten</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Alicia Keys — Girl on Fire</td>
<td>20</td>
</tr>
<tr>
<td>2. PSY — Gangnam Style</td>
<td>19</td>
</tr>
<tr>
<td>3. Florida Georgia Line — Cruise</td>
<td>23</td>
</tr>
<tr>
<td>4. Taylor Swift — I Knew You Were Trouble</td>
<td>8</td>
</tr>
<tr>
<td>5. 2 Chainz — I'm Different</td>
<td>6</td>
</tr>
<tr>
<td>6. Rihanna — Diamonds</td>
<td>15</td>
</tr>
<tr>
<td>7. Gary Allan — Every Storm</td>
<td>8</td>
</tr>
<tr>
<td>8. Little Big Town — Pontoon</td>
<td>34</td>
</tr>
<tr>
<td>9. The Band Perry — Better Dig Two</td>
<td>8</td>
</tr>
<tr>
<td>10. Lil Wayne — No Worries</td>
<td>15</td>
</tr>
</tbody>
</table>
   Source: www.billboard.com

Are these data discrete or continuous?

43. How’s the economy? A poll conducted by the American Research Group asked individuals their views on how the economy will be a year from now. Respondents were given four choices: Better than today, Same as today, Worse than today, and Undecided. Are these choices nominal or ordinal?

44. Global warming: A poll conducted in October 2012 asked people between the ages of 18 and 29 how serious a problem global warming is. Of those who responded, 43% thought it was very serious, 24% thought it was somewhat serious, 15% thought it was not too serious, and 17% thought it was not a problem. Are these percentages qualitative or quantitative?

45. Read any good books lately? According to Time magazine, some of the best fiction books in 2012 were:
   This Is How You Lose Her by Junot Diaz
   Where’d You Go, Bernadette? by Maria Semple
46. Watch your language: According to Merriam-Webster Online, the top ten Funny Sounding and Interesting words are:

1. Bumfuzzle
2. Cattywampus
3. Gardyloo
4. Taradiddle
5. Billingsgate
6. Snickersnee
7. Widdershins
8. Collywobles
9. Gubbins
10. Diphthong

47. Top ten PC games: In February 2012, Nielsen Media published the following data about the top ten PC games:

<table>
<thead>
<tr>
<th>Game Title</th>
<th>Publisher</th>
<th>Percentage of Gaming Audience</th>
<th>Average Minutes Played per Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>World of Warcraft</td>
<td>Blizzard Entertainment</td>
<td>7.154</td>
<td>500</td>
</tr>
<tr>
<td>League of Legends</td>
<td>Riot Games</td>
<td>3.776</td>
<td>402</td>
</tr>
<tr>
<td>Hanging Gardens of Babylon</td>
<td>Big Fish Games, Inc.</td>
<td>6.894</td>
<td>267</td>
</tr>
<tr>
<td>Lord of the Rings Online: The Shadows of Angmar</td>
<td>Turbine</td>
<td>1.249</td>
<td>527</td>
</tr>
<tr>
<td>Half-Life 2</td>
<td>Vivendi Games</td>
<td>2.631</td>
<td>265</td>
</tr>
<tr>
<td>Warcraft III: Reign of Chaos</td>
<td>Blizzard Entertainment</td>
<td>1.062</td>
<td>364</td>
</tr>
<tr>
<td>The Sims 3</td>
<td>The Electronic Arts Inc.</td>
<td>1.308</td>
<td>227</td>
</tr>
<tr>
<td>Zuma's Revenge!—Adventure</td>
<td>PopCap Games</td>
<td>1.444</td>
<td>148</td>
</tr>
<tr>
<td>Dungeons &amp; Dragons Online: Stormreach</td>
<td>Atari</td>
<td>0.388</td>
<td>445</td>
</tr>
<tr>
<td>Grand Theft Auto: San Andreas</td>
<td>Rockstar Games</td>
<td>0.752</td>
<td>188</td>
</tr>
</tbody>
</table>

a. Which of the columns represent qualitative variables?
b. Which of the columns represent quantitative variables?
c. Which of the columns represent nominal variables?
d. Which of the columns represent ordinal variables?

48. At the movies: The following table provides information about the top-grossing movies for the years 1997–2012:

<table>
<thead>
<tr>
<th>Year</th>
<th>Movie Title</th>
<th>Creative Type</th>
<th>MPAA Rating</th>
<th>Ticket Sales (millions)</th>
<th>Tickets Sold (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>Men in Black</td>
<td>Science fiction</td>
<td>PG-13</td>
<td>250.1</td>
<td>54.5</td>
</tr>
<tr>
<td>1998</td>
<td>Titanic</td>
<td>Dramatization</td>
<td>PG-13</td>
<td>443.3</td>
<td>94.5</td>
</tr>
<tr>
<td>1999</td>
<td>Star Wars Ep. I: The Phantom Menace</td>
<td>Science fiction</td>
<td>PG</td>
<td>430.4</td>
<td>84.7</td>
</tr>
<tr>
<td>2000</td>
<td>How the Grinch Stole Christmas</td>
<td>Kids fiction</td>
<td>PG</td>
<td>253.4</td>
<td>47.0</td>
</tr>
<tr>
<td>2001</td>
<td>Harry Potter and the Sorcerer's Stone</td>
<td>Fantasy</td>
<td>PG</td>
<td>291.6</td>
<td>51.5</td>
</tr>
<tr>
<td>2002</td>
<td>Spider-Man</td>
<td>Super hero</td>
<td>PG-13</td>
<td>403.7</td>
<td>69.5</td>
</tr>
<tr>
<td>2003</td>
<td>Finding Nemo</td>
<td>Kids fiction</td>
<td>G</td>
<td>339.7</td>
<td>56.3</td>
</tr>
<tr>
<td>2004</td>
<td>Shrek 2</td>
<td>Kids fiction</td>
<td>PG</td>
<td>436.5</td>
<td>70.3</td>
</tr>
<tr>
<td>2005</td>
<td>Star Wars Ep. III: Revenge of the Sith</td>
<td>Science fiction</td>
<td>PG-13</td>
<td>380.3</td>
<td>59.3</td>
</tr>
<tr>
<td>2006</td>
<td>Pirates of the Caribbean: Dead Man's Chest</td>
<td>Fantasy</td>
<td>PG-13</td>
<td>423.3</td>
<td>64.6</td>
</tr>
<tr>
<td>2007</td>
<td>Spider-Man 3</td>
<td>Super hero</td>
<td>PG-13</td>
<td>336.5</td>
<td>48.9</td>
</tr>
<tr>
<td>2008</td>
<td>The Dark Knight</td>
<td>Super hero</td>
<td>PG-13</td>
<td>531.0</td>
<td>74.0</td>
</tr>
<tr>
<td>2009</td>
<td>Transformers: Revenge of the Fallen</td>
<td>Science fiction</td>
<td>PG-13</td>
<td>402.1</td>
<td>53.6</td>
</tr>
<tr>
<td>2010</td>
<td>Toy Story 3</td>
<td>Kids fiction</td>
<td>G</td>
<td>415.0</td>
<td>52.6</td>
</tr>
<tr>
<td>2011</td>
<td>Harry Potter and the Deathly Hallows: Part II</td>
<td>Fantasy</td>
<td>PG-13</td>
<td>381.0</td>
<td>48.0</td>
</tr>
<tr>
<td>2012</td>
<td>Marvel's The Avengers</td>
<td>Super hero</td>
<td>PG-13</td>
<td>623.3</td>
<td>79.6</td>
</tr>
</tbody>
</table>

Source: http://www.th-numbers.com/

a. Which of the columns represent qualitative variables?
b. Which of the columns represent quantitative variables?
c. Which of the columns represent nominal variables?
d. Which of the columns represent ordinal variables?
How can we prevent confounding? One way is to design a study so that the confounder isn’t a factor. For example, to determine whether smoking increases the risk of heart attack, we could compare a group of male smokers to a group of male nonsmokers, and a group of female smokers to a group of female nonsmokers. Gender wouldn’t be a confounder here, because there would be no differences in gender between the smoking and nonsmoking groups. Of course, there are other possible confounders. Smoking rates vary among ethnic groups, and rates of heart attacks do, too. If people in ethnic groups that are more susceptible to heart attacks are also more likely to smoke, then ethnicity becomes a confounder. This can be dealt with by comparing smokers of the same gender and ethnic group to nonsmokers of that gender and ethnic group.

Designing observational studies that are relatively free of confounding is difficult. In practice, many studies must be conducted over a long period of time. In the case of smoking, this has been done, and we can be confident that smoking does indeed increase the risk of heart attack, along with other diseases. If you don’t smoke, you have a much better chance to live a long and healthy life.

**SUMMARY**

In an observational study, when there are differences in the outcomes among the treatment groups, it is often difficult to determine whether the differences are due to the treatments or to confounding.

**EXAMPLE 1.16**

**Determining the effect of confounding**

In a study of the effects of blood pressure on health, a large group of people of all ages were given regular blood pressure checkups for a period of one year. It was found that people with high blood pressure were more likely to develop cancer than people with lower blood pressure. Explain how this result might be due to confounding.

**Solution**

Age is a likely confounder. Older people tend to have higher blood pressure than younger people, and older people are more likely to get cancer than younger people. Therefore people with high blood pressure may have higher cancer rates than younger people, even though high blood pressure does not cause cancer.

**Check Your Understanding**

1. To study the effect of air pollution on respiratory health, a group of people in a city with high levels of air pollution and another group in a rural area with low levels of pollution are examined to determine their lung capacity. Is this a randomized experiment or an observational study? *Observational study*

2. It is known that drinking alcohol increases the risk of contracting liver cancer. Assume that in an observational study, a group of smokers has a higher rate of liver cancer than a group of nonsmokers. Explain how this result might be due to confounding. *Smokers may drink more than nonsmokers.*

*Answers are on page 26.*

**Objective 4** Describe various types of observational studies

**Types of Observational Studies**

There are two main types of observational studies: cohort studies and case-control studies. Cohort studies can be further divided into prospective, cross-sectional, and retrospective studies.
Case-control studies are always retrospective, because the outcome (case or control) has occurred before the sampling is done. Case-control studies have the same advantages and disadvantages as retrospective cohort studies. In addition, case-control studies have the advantage that they can be used to study rare diseases.

Check Your Understanding

3. In a study conducted at the University of Southern California, J. Peters and colleague studied elementary school students in 12 California communities. Each year for 10 years, they measured the respiratory function of the children and the levels of air pollution in the communities.
   a. Was this a cohort study or a case-control study?  Case-control
   b. Was the study prospective, cross-sectional, or retrospective?  Prospective

4. In a study conducted at the University of Colorado, J. Ruttenber and colleagues studied people who had worked at the Rocky Flats nuclear weapons production facility near Denver, Colorado. They studied a group of workers who had contracted lung cancer, and another group who had not contracted lung cancer. They looked back at plant records to determine the amount of radiation exposure for each worker.
   a. Was this a cohort study or a case-control study?  Case-control
   b. Was the study prospective, cross-sectional, or retrospective?  Retrospective

Answers are on page 21

SECTION 1.3 Exercises

Exercises 1–4 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 5–10, fill in each blank with the appropriate word or phrase.

5. In a ________ experiment, subjects do not decide for themselves which treatment they will get.  randomized

6. In a ________ study, neither the investigators nor the subjects know who is getting which treatment.  double-blind

7. A study in which the assignment to treatment groups is not made by the investigator is called ________.  observational

8. A ________ is a variable related to both the treatment and the outcome.  confounder

9. In a ________ study, the subjects are followed over time.  prospective

10. In a ________ study, a group of subjects is studied to determine whether various factors of interest are associated with an outcome.  cohort

In Exercises 11–16, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

11. In a randomized experiment, the treatment groups do not differ in any systematic way except that they receive different treatments.  True

12. A confounder makes it easier to draw conclusions from a study.  False

13. In an observational study, subjects are assigned to treatment groups at random.  False

14. Observational studies are generally more reliable than randomized experiments.  False

15. In a case-control study, the outcome has occurred before the subjects are sampled.  True

16. In a cross-sectional study, measurements are made at only one point in time.  True

Practicing the Skills

17. To determine the effectiveness of a new pain reliever, a randomly chosen group of pain sufferers is assigned to take the new drug, and another randomly chosen group is assigned to take a placebo.
   a. Is this a randomized experiment or an observational study?  Randomized experiment
   b. The subjects taking the new drug experienced substantially more pain relief than those taking the placebo. The research team concluded that the new drug is effective in relieving pain. Is this conclusion well justified? Explain.  Yes

18. A medical researcher wants to determine whether exercise can lower blood pressure. At a health fair, he measures the blood pressure of 100 individuals, and interviews them about their exercise habits. He divides the individuals into two categories: those whose typical level of exercise is low and those whose level of exercise is high.
   a. Is this a randomized experiment or an observational study?  Observational study
b. The subjects in the low-exercise group had considerably higher blood pressure, on the average, than subjects in the high-exercise group. The researcher concluded that exercise decreases blood pressure. Is this conclusion well justified? Explain. Yes

19. A medical researcher wants to determine whether exercising can lower blood pressure. She recruits 100 people with high blood pressure to participate in the study. She assigns a random sample of 50 of them to pursue an exercise program that includes daily swimming and jogging. She assigns the other 50 to refrain from vigorous activity. She measures the blood pressure of each of the 100 individuals both before and after the study.
   a. Is this a randomized experiment or an observational study? Randomized experiment
   b. On the average, the subjects in the exercise group substantially reduced their blood pressure, while the subjects in the no-exercise group did not experience a reduction. The researcher concluded that exercise decreases blood pressure. Is this conclusion well justified? Explain. Yes

20. An agricultural scientist wants to determine the effect of fertilizer type on the yield of tomatoes. There are four types of fertilizer under consideration. She plants tomatoes on four plots of land. Each plot is treated identically except for receiving a different type of fertilizer.
   a. What are the treatments? Four types of fertilizer
   b. Is this a randomized experiment or an observational study? Randomized experiment
   c. The yields differ substantially among the four plots. Can you conclude that the differences in yield are due to the differences in fertilizer? Explain. Yes

---

Working with the Concepts

21. Air pollution and colds: A scientist wants to determine whether people who live in places with high levels of air pollution get more colds than people in areas with little air pollution. Do you think it is possible to design a randomized experiment to study this question, or will an observational study be necessary? Explain. Observational study necessary

22. Cold medications: A scientist wants to determine whether a new cold medicine relieves symptoms more effectively than a currently used medicine. Do you think it is possible to design a randomized experiment to study this question, or will an observational study be necessary? Explain. Randomized experiment possible

23. Taxicabs and crime: A sociologist discovered that regions with more taxicabs tend to have higher crime rates. Does increasing the number of taxicabs cause the crime rate to increase, or could the result be due to confounding? Explain. Could be due to confounding

24. Recovering from heart attacks: In a study of people who had suffered heart attacks, it was found that those who lived in smaller houses were more likely to recover than those who lived in larger houses. Does living in a smaller house increase the likelihood of recovery from a heart attack, or could the result be due to confounding? Explain. Could be due to confounding

25. Eat your vegetables: In an observational study, people who ate four or more servings of fresh fruits and vegetables each day were less likely to develop colon cancer than people who ate little fruit or vegetables. True or false:
   a. The results of the study show that eating more fruits and vegetables reduces your risk of contracting colon cancer. True
   b. The results of the study may be due to confounding, since the lifestyles of people who eat large amounts of fruits and vegetables may differ in many ways from those of people who do not. True

26. Vocabulary and height: A vocabulary test was given to students at an elementary school. The students’ ages ranged from 5 to 11 years old. It was found that the students with larger vocabularies tended to be taller than the students with smaller vocabularies. Explain how this result might be due to confounding.

27. Secondhand smoke: A recent study compared the heart rates of 19 infants born to nonsmoking mothers with those of 17 infants born to mothers who smoked an average of 15 cigarettes a day while pregnant and after giving birth. The heart rates of the infants at one year of age were 2% lower on the average for the smoking mothers. Does the heart rate of infants born to smoking mothers differ at one year of age? Yes
   a. What is the outcome variable? Heart rate
   b. What is the treatment variable? Maternal smoking
   c. Was this a cohort study or a case-control study? Cohort
   d. Was the study prospective, cross-sectional, or retrospective? Prospective
   e. Could the results be due to confounding? Explain. Yes

Source: Environmental Health Perspectives 118:158–159

28. Pollution in China: In a recent study, Z. Zhao and colleagues measured the levels of formaldehyde in the air in 34 classrooms in the schools in the city of Taiyuan, China. On the same day, they gave questionnaires to 1993 students aged 11–15 in those schools, asking them whether they had experienced respiratory problems (such as asthma attacks, wheezing, or shortness of breath). They found that the students in the classrooms with higher levels of formaldehyde reported more respiratory problems. Does an increase in formaldehyde levels cause an increase in respiratory problems? Explain. No
   a. What is the outcome variable? Respiratory problems
   b. What is the treatment variable? Formaldehyde level
   c. Was this a cohort study or a case-control study? Cohort
   d. Was the study prospective, cross-sectional, or retrospective? Cross-sectional
   e. Could the results be due to confounding? Explain. Yes

Source: Environmental Health Perspectives 116:90–97
Extending the Concepts

29. The Salk Vaccine Trial: In 1954, the first vaccine against polio, known as the Salk vaccine, was tested in a large randomized double-blind study. Approximately 750,000 children were asked to enroll in the study. Of these, approximately 350,000 did not participate, because their parents refused permission. The children who did participate were randomly divided into two groups of about 200,000 each. One group, the treatment group, got the vaccine, while the other group, the control group, got a placebo. The rate of polio in the treatment group was less than half of that in the control group.
   a. Is it reasonable to conclude that the Salk vaccine was effective in reducing the rate of polio?  Yes
   b. Polio is sometimes difficult to diagnose, as its early symptoms are similar to those of the flu. Explain why it was important for the doctors in the study not to know which children were getting the vaccine.
   c. Perhaps surprisingly, polio was more common among upper-income and middle-income children than among lower-income children. The reason is that lower-income children tended to live in less hygienic surroundings. They would contract mild cases of polio in infancy while still protected by their mother’s antibodies, and thereby develop a resistance to the disease. The children who did not participate in the study were more likely to come from lower-income families. The rate of polio in this group was substantially lower than the rate in the placebo group. Does this prove that the placebo caused polio, or could this be due to confounding? Explain. Could be due to confounding.

30. Another Salk Vaccine Trial: Another study of the Salk vaccine, conducted at the same time as the trial described in Exercise 29, used a different design. In this study, approximately 350,000 second graders were invited to participate. About 225,000 did so, and the other 125,000 refused. All of the participating second graders received the vaccine. The control group consisted of approximately 725,000 first and third graders. They were not given any placebo, so no consent was necessary.
   a. Was this a randomized experiment?  No
   b. Was it double-blind?  No
   c. The treatment group consisted of children who had consent to participate. The control group consisted of all first and third graders. It turned out that the results of this study seriously underestimated the effectiveness of the vaccine. Use the information provided in Exercise 29(c) to explain why.

Answers to Check Your Understanding Exercises for Section 1.3

1. Observational study
2. People who smoke may be more likely to drink alcohol than people who do not smoke. Therefore, it might be possible for smokers to have higher rates of liver cancer without it being caused by smoking.
3. a. Cohort study  b. Prospective
4. a. Case-control study  b. Retrospective

SECTION 1.4 Bias in Studies

Objectives
1. Define bias
2. Identify sources of bias

Defining Bias
No study is perfect, and even a properly conducted study will generally not give results that are exactly correct. For example, imagine that you were to draw a simple random sample of students at a certain college to estimate the percentage of students who are Democrats. Your sample would probably contain a somewhat larger or smaller percentage of Democrats than the entire population of students, just by chance. However, imagine drawing many simple random samples. Some would have a greater percentage of Democrats than in the population, and some would have a smaller percentage of Democrats than in the population. But on the average, the percentage of Democrats in a simple random sample will be the same as the percentage in the population. A study conducted by a procedure that produces the correct result on the average is said to be unbiased.

Now imagine that you tried to estimate the percentage of Democrats in the population by selecting students who attended a speech made by a Democratic politician. On
The words “heavy” and “burden” suggest that taxes are too high, and encourage a “Yes” response. A better way to ask this question is to present it as a multiple choice:

“What is your opinion on decreasing taxes for middle-class families?
Choices: Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree.”

Nonresponse bias
People cannot be forced to answer questions or to participate in a study. In any study, a certain proportion of people who are asked to participate refuse to do so. These people are called nonresponders. In many cases, the opinions of nonresponders tend to differ from the opinions of the ones who do respond. As a result, surveys with many nonresponders are often biased.

Sampling bias
Sampling bias occurs when some members of the population are more likely to be included in the sample than others. For example, samples of convenience almost always have sampling bias, because people who are easy to sample are more likely to be included. It is almost impossible to avoid sampling bias completely, but modern survey organizations work hard to keep it at a minimum.

A Big Sample Size Doesn’t Make Up for Bias
A sample is useful only if it is drawn by a method that is likely to represent the population well. If you use a biased method to draw a sample, then drawing a big sample doesn’t help; a big nonrepresentative sample does not describe a population any better than a small nonrepresentative sample. In particular, voluntary response surveys often draw several hundred thousand people to participate. Although the sample is large, it is unlikely to represent the population well, so the results are meaningless.

Check Your Understanding

1. Eighty thousand people attending a professional football game filled out surveys asking their opinions on using tax money to upgrade the football stadium. Seventy percent said that they supported the use of tax money. Then a pollster surveyed a simple random sample of 500 voters, and only 30% of the voters in this sample supported the use of tax money. The owner of the football team claims that the survey done at the football stadium is more reliable, because the sample size was much larger. Is he right? Explain. No

2. A polling organization placed telephone calls to 1000 people in a certain city to ask them whether they favor a tax increase to build a new school. Two hundred people answered the phone, and 150 of them opposed the tax. Can you conclude that a majority of people in the city oppose the tax, or is it likely that this result is due to bias? Explain. Likely due to bias

4. People who are asked to participate in a study but refuse to do so are called ________. nonresponders

5. A large sample is useful only if it is drawn by a method that is likely to represent the ________. population

In Exercises 6–8, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

6. The way that a question in a survey is worded rarely has an effect on the responses. False

SECTION 1.4 Exercises

Exercises 1 and 2 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 3–5, fill in each blank with the appropriate word or phrase.

3. ________ are highly unreliable in part because people who have strong opinions are more likely to participate. Voluntary response surveys

4. People who are asked to participate in a study but refuse to do so are called ________. nonresponders

5. A large sample is useful only if it is drawn by a method that is likely to represent the ________. population

In Exercises 6–8, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

6. The way that a question in a survey is worded rarely has an effect on the responses. False
7. Surveys with many nonresponders often provide misleading results. **True**

8. Large samples usually give reasonably accurate results, no matter how they are drawn. **False**

**Practicing the Skills**

In Exercises 9–16, specify the type of bias involved.

9. A bank sent out questionnaires to a simple random sample of 500 customers asking whether they would like the bank to extend its hours. Eighty percent of those returning the questionnaire said they would like the bank to extend its hours. Of the 500 questionnaires, 20 were returned. **Nonresponse**

10. To determine his constituents’ feelings about election reform, a politician sends a survey to people who have subscribed to his newsletter. More than 1000 responses are received. **Voluntary response**

11. An e-store that sells cell phone accessories reports that 98% of its customers are satisfied with the speed of delivery. **Self-interest**

12. A sign in a restaurant claims that 95% of their customers believe them to have the best food in the world. **Self-interest**

13. A television newscaster invites viewers to email their opinions about whether the U.S. Congress is doing a good job in handling the economy. More than 100,000 people send an opinion. **Voluntary response**

14. A police department conducted a survey in which police officers interviewed members of their community to ask their opinions on the effectiveness of the police department. The police chief reported that 90% of the people interviewed said that they were satisfied with the performance of the police department. **Social acceptability**

15. In a study of the effectiveness of wearing seat belts, a group of people who had survived car accidents in which they had not worn seat belts reported that seat belts would not have helped them. **Nonresponse**

16. To estimate the prevalence of illegal drug use in a certain high school, the principal interviewed a simple random sample of 100 students and asked them about their drug use. Five percent of the students acknowledged using illegal drugs. **Social acceptability**

**Working with the Concepts**

17. **Nuclear power, anyone?** In a survey conducted by representatives of the nuclear power industry, people were asked the question: “Do you favor the construction of nuclear power plants in order to reduce our dependence on foreign oil?” A group opposed to the use of nuclear power conducted a survey with the question: “Do you favor the construction of nuclear power plants that can kill thousands of people in an accident?”

   a. Do you think that the percentage of people favoring the construction of nuclear power plants would be about the same in both surveys? **No**

   b. Would either of the two surveys produce reliable results? **Explain. No**

18. **Who’s calling, please?** Random-digit dialing is a sampling method in which a computer generates phone numbers at random to call. In recent years, caller ID has become popular. Do you think that caller ID increases the bias in random digit dialing? **Explain. Yes**

19. **Who’s calling, please?** Many polls are conducted over the telephone. Some polling organizations choose a sample of phone numbers to call from lists that include landline phone numbers only, and do not include cell phones. Do you think this increases the bias in phone polls? **Explain. Yes**

20. **Order of choices:** When multiple-choice questions are asked, the order of the choices is usually changed each time the question is asked. For example, in the 2012 presidential election, a pollster would ask one person “Who do you prefer for president, Barack Obama or Mitt Romney?” For the next person, the order of the names would be reversed: “Mitt Romney or Barack Obama?” If the choices were given in the same order each time, do you think that might introduce bias? **Explain. Yes**

**Extending the Concepts**

21. **Literary Digest poll:** In the 1936 presidential election, Republican candidate Alf Landon challenged President Franklin Roosevelt. The *Literary Digest* magazine conducted a poll in which they mailed questionnaires to more than 10 million voters. The people who received the questionnaires were drawn from lists of automobile owners and people with telephones. The magazine received 2.3 million responses, and predicted that Landon would win the election in a landslide with 57% of the vote. In fact, Roosevelt won in a landslide with 62% of the vote. Soon afterward, the *Literary Digest* folded.

   a. In 1936 most people did not own automobiles, and many did not have telephones. **Explain how this could have caused the results of the poll to be mistaken.**

   b. What can be said about the response rate? **Explain how this could have caused the results of the poll to be mistaken.**

   c. The *Literary Digest* believed that its poll would be accurate, because it received 2.3 million responses, which is a very large number. **Explain how the poll could be wrong, even with such a large sample.**

**Answers to Check Your Understanding Exercises for Section 1.4**

1. No. The sample taken at the football stadium is biased, because football fans are more likely to be sampled than others. The fact that the sample is big doesn’t make it any better.

2. No. There is a high degree of nonresponse bias in this sample.
Chapter 1 Summary

Section 1.1: Most populations are too large to allow us to study each member, so we draw samples and study those. Samples must be drawn by an appropriate method. Simple random sampling, stratified sampling, cluster sampling, and systematic sampling are all valid methods. When none of these methods are feasible, a sample of convenience may be used, so long as it is reasonable to believe that there is no systematic difference between the sample and the population.

Section 1.2: Data sets contain values of variables. Qualitative variables place items in categories, whereas quantitative variables count or measure. Qualitative variables can be either ordinal or nominal. An ordinal variable is one for which the categories have a natural ordering. For nominal variables, the categories have no natural ordering. Quantitative variables can be discrete or continuous. Discrete variables are ones whose possible values can be listed, whereas continuous variables can take on any value anywhere within an interval.

Section 1.3: Scientists conduct studies to determine whether different treatments produce different outcomes. The most reliable studies are randomized experiments, in which subjects are assigned to treatments at random. When randomized experiments are not feasible observational studies may be performed. Results of observational studies may be hard to interpret, because of the potential for confounding.

Section 1.4: Some studies produce more reliable results than others. A study that is conducted by a method that tends to produce an incorrect result is said to be biased. Some of the most common forms of bias are voluntary response bias, self-interest bias, social acceptability bias, leading question bias, nonresponse bias, and sampling bias.

Vocabulary and Notation

<table>
<thead>
<tr>
<th>Bias</th>
<th>Nominal Variable</th>
<th>Sampling Bias</th>
<th>Seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>Nonresponders</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>Biased</td>
<td>Nonresponse bias</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>Case-control study</td>
<td>23</td>
<td>Observational study</td>
<td>19</td>
</tr>
<tr>
<td>Categorical variable</td>
<td>13</td>
<td>Ordinal variable</td>
<td>13</td>
</tr>
<tr>
<td>Cluster sampling</td>
<td>6</td>
<td>Outcome</td>
<td>19</td>
</tr>
<tr>
<td>Cohort study</td>
<td>23</td>
<td>Parameter</td>
<td>8</td>
</tr>
<tr>
<td>Completely randomized experiment</td>
<td>21</td>
<td>Population</td>
<td>2</td>
</tr>
<tr>
<td>Confounder</td>
<td>21</td>
<td>Prospective study</td>
<td>23</td>
</tr>
<tr>
<td>Confounding</td>
<td>21</td>
<td>Qualitative variable</td>
<td>13</td>
</tr>
<tr>
<td>Continuous variable</td>
<td>14</td>
<td>Quantitative variable</td>
<td>13</td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td>23</td>
<td>Randomized block experiment</td>
<td>21</td>
</tr>
<tr>
<td>Data</td>
<td>12</td>
<td>Randomized experiment</td>
<td>19</td>
</tr>
<tr>
<td>Data set</td>
<td>12</td>
<td>Response</td>
<td>19</td>
</tr>
<tr>
<td>Discrete variable</td>
<td>14</td>
<td>Retrospective study</td>
<td>23</td>
</tr>
<tr>
<td>Double-blind</td>
<td>20</td>
<td>Sample</td>
<td>2</td>
</tr>
<tr>
<td>Experimental unit</td>
<td>18</td>
<td>Sample of convenience</td>
<td>5</td>
</tr>
<tr>
<td>Individual</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leading question bias</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter Quiz

1. Provide an example of a qualitative variable and an example of a quantitative variable.
2. Is the name of your favorite author a qualitative variable or a quantitative variable? Qualitative
3. True or false: Nominal variables do not have a natural ordering. True
4. __________ variables are quantitative variables that can take on any value in some interval. Continuous
5. True or false: Ideally, a sample should represent the population as little as possible. False
6. A utility company sends surveys to 200 of its customers in such a way that 100 surveys are sent to customers who pay their bills on time, 50 surveys are sent to customers whose bills are less than 30 days late, and 50 surveys are sent to customers whose bills are more than 30 days late. Which type of sample does this represent? Stratified sample
7. A sample of convenience is ________, when it is reasonable to believe that there is no systematic difference between the sample and the population. (Choices: acceptable, not acceptable) Acceptable
8. The manager of a restaurant walks around and asks selected customers about the service they have received. Which type of sample does this represent? Sample of convenience
9. True or false: An experiment where neither the investigators nor the subjects know who has been assigned to which treatment is called a double-blind experiment. True
10. A poll is conducted of 3500 households close to major national airports, and another 2000 that are not close to an airport, in order to study whether living in a noisier environment results in health effects. Is this a randomized experiment or an observational study?  
Observational study

11. In a study, 200 patients with skin cancer are randomly divided into two groups. One group receives an experimental drug and the other group receives a placebo. Is this a randomized experiment or an observational study?  
Randomized experiment

12. In a randomized experiment, if there are large differences in outcomes among treatment groups, we can conclude that the differences are due to the _________.  
Differences in treatment

13. In analyzing the course grades of students in an elementary statistics course, a professor notices that students who are seniors performed better than students who are sophomores. The professor is tempted to conclude that older students perform better than younger ones. Describe a possible confounder in this situation.  
Seniors may be better prepared

14. True or false: The way that questions are worded on a survey may have an effect on the responses.  
True

15. A radio talk show host invites listeners to call the show to express their opinions about a political issue. How reliable is this survey? Explain.  
Not reliable

### Review Exercises

1. Qualitative or quantitative? Is the number of points scored in a football game qualitative or quantitative?  
Quantitative

2. Nominal or ordinal? Is the color of an MP3 player nominal or ordinal?  
Nominal

3. Discrete or continuous? Is the area of a college campus discrete or continuous?  
Continuous

4. Which type of variable is it? A theater concession stand sells soft drink and popcorn combos that come in sizes small, medium, large, and jumbo. True or false:  
a. Size is a qualitative variable.  
True

b. Size is an ordinal variable.  
True

c. Size is a continuous variable.  
False

In Exercises 5–8, identify the kind of sample that is described.

5. Website ratings: A popular website is interested in conducting a survey of 400 visitors to the site in such a way that 200 of them will be under age 30, 150 will be aged 30–55, and 50 will be over 55.  
Stratified sample

6. Favorite performer: Viewers of a television show are asked to vote for their favorite performer by sending a text message to the show.  
Voluntary response sample

7. School days: A researcher selects a random sample of 12 high schools in a certain region and surveys all of the administrative staff members in each school about a potential change in the ordering of supplies. Which type of sample does this represent?  
Cluster sample

8. Political polling: A pollster obtains a list of registered voters and uses a computer random number generator to choose 100 of them to ask which candidate they prefer in an upcoming election.  
Simple random sample

9. Fluoride and tooth decay: Researchers examine the association between the fluoridation of water and the prevention of tooth decay by comparing the prevalence of tooth decay in countries that have fluoridated water with the prevalence in countries that do not.  
a. Is this a randomized experiment or an observational study?  
Observational study

b. Assume that tooth decay was seen to be less common in countries with fluoridated water. Could this result be due to confounding? Explain.  
Yes

10. Better gas mileage: A taxi company in a large city puts a new type of tire on a random sample of 50 cars, and the regular type of tire on another random sample of 50 cars. After a month, the gas mileage of each car was measured.  
a. Is this a randomized experiment or an observational study?  
Randomized experiment

b. Assume that one of the samples had noticeably better gas mileage than the other. Could this result be due to confounding? Explain.  
Yes

11. Cell phones and driving: To determine whether using a cell phone while driving increases the risk of an accident, a researcher examines accident reports to obtain data about the number of accidents in which a driver was talking on a cell phone.  
a. Is this a randomized experiment or an observational study?  
Observational study

b. Assume that the accident reports show that people were more likely to have an accident while talking on a cell phone. Could this result be due to confounding? Explain.  
Yes

12. Turn in your homework: The English department at a local college is considering using electronic-based assignment submission in its English composition classes. To study its effects, each section of the class is divided into two groups at random. In one group, assignments are submitted by turning them in to the professor on paper. In the other group, assignments are submitted electronically.  
a. Is this a randomized experiment or an observational study?  
Randomized experiment

b. Assume that the electronically submitted assignments had many fewer typographical errors, on average, than the ones submitted on paper. Could this result be due to confounding? Explain.  
Yes
In Exercises 13–15, explain why the results of the studies described are unreliable.

13. Which TV station do you watch? The TV columnist for a local newspaper invites readers to log on to a website to vote for their favorite TV newscaster. Voluntary response sample

14. Longevity: A life insurance company wants to study the life expectancy of people born in 1950. The company’s actuaries examine death certificates of people born in that year to determine how long they lived. Nonresponse bias

15. Political opinion: A congressman sent out questionnaires to 10,000 constituents to ask their opinions on a new health-care proposal. A total of 200 questionnaires were returned, and 70% of those responding supported the proposal. Nonresponse bias

Write About It

1. Describe the difference between a stratified sample and a cluster sample.
2. Explain why it is better, when possible, to draw a simple random sample rather than a sample of convenience.
3. Describe circumstances under which each of the following samples could be used: simple random sample, a sample of convenience, a stratified sample, a cluster sample, a systematic sample.
4. Suppose that you were asked to collect some information about students in your class for a statistics project. Give some examples of variables you might collect that are ordinal, nominal, discrete, and continuous.
5. Quantitative variables are numerical. Are some qualitative variables numerical as well? If not, explain why not. If so, provide an example.
6. What are the primary differences between a randomized experiment and an observational study?
7. What are the advantages of a double-blind study? Are there any disadvantages?
8. Provide an example of a study, either real or hypothetical, that is conducted by people who have an interest in the outcome. Explain how the results might possibly be misleading.
9. Explain why each of the following questions is leading. Provide a more appropriate wording.
   a. Should Americans save more money or continue their wasteful spending?
   b. Do you support more funding for reputable organizations like the Red Cross?

Case Study: Air Pollution And Respiratory Symptoms

Air pollution is a serious problem in many places. One form of air pollution that is suspected to cause respiratory illness is particulate matter (PM), which consists of tiny particles in the air. Particulate matter can come from many sources, most commonly ash from burning, but also from other sources such as tiny particles of rubber that wear off of automobile and truck tires.

The town of Libby, Montana, was recently the focus of a study on the effect of PM on the respiratory health of children. Many houses in Libby are heated by wood stoves, which produce a lot of particulate pollution. The level of PM is greatest in the winter when more stoves are being used, and declines as the weather becomes warmer. The study attempted to determine whether higher levels of PM affect the respiratory health of children. In one part of the study, schoolchildren were given a questionnaire to bring home to their parents. Among other things, the questionnaire asked whether the child had experienced symptoms of wheezing during the past 60 days.

Most parents returned the questionnaire within a couple of weeks. Parents who did not respond promptly were sent another copy of the questionnaire through the mail. Many of these parents responded to this mailed version.

Table 1.2 presents, for each day, the number of questionnaires that were returned by parents of children who wheezed, the number returned by those who did not wheeze, the average concentration of particulate matter in the atmosphere during the past 60 days (in units of micrograms per cubic meter), and whether the questionnaires were delivered in school or through the mail.

We will consider a PM level of 17 or more to be high exposure, and a PM level of less than 17 to be low exposure.

1. How many people had high exposure to PM? 450
2. How many of the high-exposure people had wheeze symptoms? 41
3. What percentage of the high-exposure people had wheeze symptoms? 9.7%
4. How many people had low exposure to PM? 43
5. How many of the low-exposure people had wheeze symptoms? 2
6. What percentage of the low-exposure people had wheeze symptoms? 4.7%
7. Is there a large difference between the percentage of high-exposure people with wheeze symptoms and the percentage of low-exposure people with wheeze symptoms? Yes
8. Explain why the percentage of high-exposure people with wheeze symptoms is the same as the percentage of school-return people with wheeze symptoms.