A primary goal of statistical studies is to collect data that can then be used to make informed decisions. It should come as no surprise that the ability to make good decisions depends on the quality of the information available. The data collection step is critical to obtaining reliable information; both the type of analysis that is appropriate and the conclusions that can be drawn depend on how the data are collected. In this chapter, we first consider two types of statistical studies and then focus on widely used methods of data collection: sampling experimentation.
Chapter 2
Collecting Data Sensibly

On September 25, 2009, results from a study of the relationship between spanking and IQ were reported by a number of different news media. Some of the headlines that appeared that day were:

“Spanking lowers a child’s IQ” (Los Angeles Times)
"Do you spank? Studies indicate it could lower your kid’s IQ" (SciGuy, Houston Chronicle)
“Spanking can lower IQ” (NBC4i, Columbus, Ohio)
“Smacking hits kids’ IQ” (newscientist.com)

In the study that these headlines refer to, the investigators followed 806 kids age 2 to 4 and 704 kids age 5 to 9 for 4 years. IQ was measured at the beginning of the study and again 4 years later. The researchers found that at the end of the study, the average IQ of kids who were not spanked was 5 points higher than that of kids who were spanked among the kids who were 2 to 4 years old when the study began, and 2.8 points higher among the kids who were 5 to 9 years old when the study began.

These headlines all imply that spanking was the cause of the observed difference in IQ. Is this conclusion reasonable? The answer depends in a critical way on the study design. We’ll return to these headlines and decide if they are on target after first considering some important aspects of study design.

Observation and Experimentation

Data collection is an important step in the data analysis process. When we set out to collect information, it is important to keep in mind the questions we hope to answer on the basis of the resulting data. Sometimes we are interested in answering questions about characteristics of a single existing population or in comparing two or more well-defined populations. To accomplish this, we select a sample from each population under consideration and use the sample information to gain insight into characteristics of those populations.

For example, an ecologist might be interested in estimating the average shell thickness of bald eagle eggs. A social scientist studying a rural community may want to determine whether gender and attitude toward abortion are related. These are examples of studies that are observational in nature. In these studies, we want to observe characteristics of members of an existing population or of several populations, and then use the resulting information to draw conclusions. In an observational study, it is important to obtain a sample that is representative of the corresponding population.

Sometimes the questions we are trying to answer deal with the effect of certain explanatory variables on some response and cannot be answered using data from an observational study. Such questions are often of the form, “What happens when ...?” or, “What is the effect of ...?” For example, an educator may wonder what would happen to test scores if the required lab time for a chemistry course were increased from 3 hours to 6 hours per week. To answer such questions, the researcher conducts an experiment to collect relevant data. The value of some response variable (test score in the chemistry example) is recorded under different experimental conditions (3-hour lab and 6-hour lab). In an experiment, the researcher manipulates one or more explanatory variables, also sometimes called factors, to create the experimental conditions.
**DEFINITION**

A study is an **observational study** if the investigator observes characteristics of a sample selected from one or more existing populations. The goal of an observational study is usually to draw conclusions about the corresponding population or about differences between two or more populations. In a well-designed observational study, the sample is selected in a way that is designed to produce a sample that is representative of the population.

A study is an **experiment** if the investigator observes how a response variable behaves when one or more explanatory variables, also called factors, are manipulated. The usual goal of an experiment is to determine the effect of the manipulated explanatory variables (factors) on the response variable. In a well-designed experiment, the composition of the groups that will be exposed to different experimental conditions is determined by random assignment.

The type of conclusion that can be drawn from a statistical study depends on the study design. Both observational studies and experiments can be used to compare groups, but in an experiment the researcher controls who is in which group, whereas this is not the case in an observational study. This seemingly small difference is critical when it comes to drawing conclusions based on data from study.

A well-designed experiment can result in data that provide evidence for a cause-and-effect relationship. This is an important difference between an observational study and an experiment. In an observational study, it is impossible to draw cause-and-effect conclusions because we cannot rule out the possibility that the served effect is due to some variable other than the explanatory variable being studied. Such variables are called confounding variables.

**DEFINITION**

A **confounding variable** is one that is related to both group membership and the response variable of interest in a research study.

Consider the role of confounding variables in the following three studies:

- The article “Panel Can’t Determine the Value of Daily Vitamins” (*San Luis Obispo Tribune*, July 1, 2003) summarized the conclusions of a government advisory panel that investigated the benefits of vitamin use. The panel looked at a large number of studies on vitamin use and concluded that the results were inadequate or conflicting. A major concern was that many of the studies were observational in nature and the panel worried that people who take vitamins might be healthier just because they tend to take better care of themselves. A potential confounding variable prevented the panel from concluding that taking vitamins was the cause of observed better health among those who take vitamins.

- Studies have shown that people over age 65 who get a flu shot are less likely than those who do not to get a flu shot to die from a flu-related illness during the following year. However, recent research has shown that people over age 65 who get a flu shot are also less likely than those who don’t to die from any cause during following year (*International Journal of Epidemiology*, December 21, 2005).

**AP* EXAM TIP**

These definitions are important when deciding what conclusions can be drawn from a statistical study. If asked “What conclusions can be made from this study?”, the very first consideration should be whether the study is observational or experimental.

**AP* EXAM TIP**

Confounding is a very important concept. You should be able to recognize confounding variables given the description of a study. You should also be able to speculate intelligently about possible confounding variables when designing studies (especially experiments).
This has lead to the speculation that those over age 65 who get flu shots are healthier as a group than those who do not get flu shots. If this is the case, observational studies that compare two groups—those who get flu shots and those who do not—may overestimate the effectiveness of the flu vaccine because general health differs in the two groups. General health is a possible confounding variable in such studies.

- The article “Heartfelt Thanks to Fido” (San Luis Obispo Tribune, July 5, 2003) summarized a study that appeared in the American Journal of Cardiology (March 15, 2003). In this study researchers measured heart rate variability (a measure of the heart’s ability to handle stress) in patients who had recovered from a heart attack. They found that heart rate variability was higher (which is good and means the heart can handle stress better) for those who owned a dog than for those who did not. Should someone who suffers a heart attack immediately go out and get a dog? Well, maybe not yet. The American Heart Association recommends additional studies to determine if the improved heart rate variability is attributable to dog ownership or due to the fact that dog owners get more exercise. If in fact dog owners do tend to get more exercise than nonowners, level of exercise is a confounding variable that would prevent us from concluding that owning a dog is the cause of improved heart rate variability.

Each of the three studies described above illustrates why potential confounding variables make it unreasonable to draw a cause-and-effect conclusion from an observational study.

Let’s return to the study on spanking and IQ described at the beginning of this section. Is this study an observational study or an experiment? Two groups were compared (children who were spanked and children who were not spanked), but the researchers did not randomly assign children to the spanking or no-spanking groups. The study is observational, and so cause-and-effect conclusions such as “spanking lowers IQ” are not justified based on the observed data. What we can say is that there is evidence that, as a group, children who are spanked tend to have a lower IQ than children who are not spanked. What we cannot say is that spanking is the cause of the lower IQ. It is possible that other variables—such as home or school environment, socio-economic status, or parents’ education—are related to both IQ and whether or not a child was spanked. These are examples of possible confounding variables.

Fortunately, not everyone made the same mistake as the writers of the headlines given earlier in this section. Some examples of headlines that got it right are:

“Lower IQ’s measured in spanked children” (world-science.net)
“Children who get spanked have lower IQs” (livescience.com)
“Research suggests an association between spanking and lower IQ in children” (CBSnews.com)

Drawing Conclusions from Statistical Studies

In this section, two different types of conclusions have been described. One type involves generalizing from what we have seen in a sample to some larger population, and the other involves reaching a cause-and-effect conclusion about the effect of an explanatory variable on a response. When is it reasonable to draw such conclusions? The answer depends on the way that the data were collected. Table 2.1 summarizes the types of conclusions that can be made with different study designs.

As you can see from Table 2.1, it is important to think carefully about the objectives of a statistical study before planning how the data will be collected. Both
observational studies and experiments must be carefully designed if the resulting data are to be useful. The common sampling procedures used in observational studies are considered in Section 2.2. In Sections 2.3 and 2.4, we consider experimentation and explore what constitutes good practice in the design of simple experiments.

<table>
<thead>
<tr>
<th>Study Description</th>
<th>Reasonable to Generalize Conclusions about Group Characteristics to the Population?</th>
<th>Reasonable to Draw Cause-and-Effect Conclusion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational study with sample selected at random from population of interest</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observational study based on convenience or voluntary response sample (poorly designed sampling plan)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Experiment with groups formed by random assignment of individuals or objects to experimental conditions</td>
<td>Individuals or objects used in study are volunteers or not randomly selected from some population of interest</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Individuals or objects used in study are randomly selected from some population of interest</td>
<td>Yes</td>
</tr>
<tr>
<td>Experiment with groups not formed by random assignment to experimental conditions (poorly designed experiment)</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**TABLE 2.1 Drawing Conclusions from Statistical Studies**

**EXERCISES 2.1 - 2.12**

2.1 ✪ The article "Television’s Value to Kids: It’s All in How They Use It" (*Seattle Times*, July 6, 2005) described a study in which researchers analyzed standardized test results and television viewing habits of 1700 children. They found that children who averaged more than 2 hours of television viewing per day when they were younger than 3 tended to score lower on measures of reading ability and short-term memory.

a. Is the study described an observational study or an experiment?

b. Is it reasonable to conclude that watching two or more hours of television is the cause of lower reading scores? Explain.

2.2 ✪ The article "Acupuncture for Bad Backs: Even Sham Therapy Works" (*Time*, May 12, 2009) summarized a study conducted by researchers at the Group Health Center for Health Studies in Seattle. In this study, 638 adults with back pain were randomly assigned to one of four groups. People in group 1 received the usual care for back pain. People in group 2 received acupuncture at a set of points tailored specifically for each individual. People in group 3 received acupuncture at a standard set of points typically used in the treatment of back pain. Those in group 4 received fake acupuncture—they were poked with a toothpick at the same set of points used in the treatment of back pain. Two notable conclusions from the study were: (1) patients receiving real or fake acupuncture...
experienced a greater reduction in pain than those receiving usual care; and (2) there was no significant difference in pain reduction for those who received acupuncture (at individualized or the standard set of points) and those who received fake acupuncture toothpick pokes.

2.3 The article “Display of Health Risk Behaviors on MySpace by Adolescents” (Archives of Pediatrics and Adolescent Medicine [2009]:27–34) described a study in which researchers looked at a random sample of 500 publicly accessible MySpace web profiles posted by 18-year-olds. The content of each profile was analyzed. One of the conclusions reported was that displaying sport or hobby involvement was associated with decreased references to risky behavior (sexual references or references to substance abuse or violence).

a. Is the study described an observational study or an experiment? Explain.

b. Is it reasonable to generalize the stated conclusion to all 18-year-olds with a publicly accessible MySpace web profile? What aspect of the study supports your answer?

c. Not all MySpace users have a publicly accessible profile. Is it reasonable to generalize the stated conclusion to all 18-year-old MySpace users? Explain.

d. Is it reasonable to generalize the stated conclusion to all MySpace users with a publicly accessible profile? Explain.

2.4 Can choosing the right music make wine taste better? This question was investigated by a researcher at a university in Edinburgh (www.decanter.com/news). Each of 250 volunteers was assigned at random to one of five rooms where they were asked to taste and rate a glass of wine. In one of the rooms, no music was playing and a different style of music was playing in each of the other four rooms. The researchers concluded that cabernet sauvignon is perceived as being richer and more robust when bold music is played than when no music is heard.

a. Is the study described an observational study or an experiment?

b. Can a case be made for the researcher’s conclusion that the music played was the cause of the higher rating? Explain.

2.5 Consider the following graphical display that appeared in the New York Times:

![Graphical display](Image not available due to copyright restrictions)

Based on the data summarized in the graph, we can see that students who have a high school GPA or 3.5 or higher and a combined SAT score of over 1200 have an 89% graduation rate when they attend a “most selective” college, but only a 59% graduation rate when they attend a “least selective” college. Give an example of a potential confounding variable that might explain why the following statement is not reasonable: If all the students that have a GPA of 3.5 or higher and a combined SAT score of 1200 or higher and that were admitted to a “least selective” college were moved to a “most selective” college, the graduation rate for these students would be approximately 89%.

2.6 “Fruit Juice May Be Fueling Pudgy Preschoolers, Study Says” is the title of an article that appeared in the San Luis Obispo Tribune (February 27, 2005). This article describes a study that found that for 3- and 4-year-olds, drinking something sweet once or twice a day doubled the risk of being seriously overweight one year later. The authors of the study state:

> Total energy may be a confounder if consumption of sweet drinks is a marker for other dietary factors associated with overweight (Pediatrics, November 2005).

Give an example of a dietary factor that might be one of the potentially confounding variables the study authors are worried about.

2.7 The article “Americans are ‘Getting the Wrong Idea’ on Alcohol and Health” (Associated Press, April 19, 2005) reported that observational studies in recent years that have concluded that moderate drinking is associated with a reduction in the risk of heart disease may be misleading. The article refers to a study conducted by...
the Centers for Disease Control and Prevention that showed that moderate drinkers, as a group, tended to be better educated, wealthier, and more active than non-drinkers. Explain why the existence of these potentially confounding variables prevents drawing the conclusion that moderate drinking is the cause of reduced risk of heart disease.

2.8 An article titled “Guard Your Kids Against Allergies: Get Them a Pet” (San Luis Obispo Tribune, August 28, 2002) described a study that led researchers to conclude that “babies raised with two or more animals were about half as likely to have allergies by the time they turned six.”

a. Do you think this study was an observational study or an experiment? Explain.

b. Describe a potential confounding variable that illustrates why it is unreasonable to conclude that being raised with two or more animals is the cause of the observed lower allergy rate.

2.9 Researchers at the Hospital for Sick Children in Toronto compared babies born to mothers with diabetes to babies born to mothers without diabetes (“Conditioning and Hyperanalgesia in Newborns Exposed to Repeated Heel Lancets,” Journal of the American Medical Association [2002]: 857–861). Babies born to mothers with diabetes have their heels pricked numerous times during the first 36 hours of life in order to obtain blood samples to monitor blood sugar level. The researchers noted that the babies born to diabetic mothers were more likely to grimace or cry when having blood drawn than the babies born to mothers without diabetes. This led the researchers to conclude that babies who experience pain early in life become highly sensitive to pain. Comment on the appropriateness of this conclusion.

2.10 Based on a survey conducted on the DietSmart.com web site, investigators concluded that women who regularly watched Oprah were only one-seventh as likely to crave fattening foods as those who watched other daytime talk shows (San Luis Obispo Tribune, October 14, 2003).

a. Is it reasonable to conclude that watching Oprah caused a decrease in cravings for fattening foods? Explain.

b. Is it reasonable to generalize the results of this study to all women in the United States? To all women who watch daytime talk shows? Explain why or why not.

2.11 A survey of affluent Americans (those with incomes of $75,000 or more) indicated that 57% would rather have more time than more money (USA Today, January 29, 2003).

a. What condition on how the data were collected would make the generalization from the sample to the population of affluent Americans reasonable?

b. Would it be reasonable to generalize from the sample to say that 57% of all Americans would rather have more time than more money? Explain.

2.12 Does living in the South cause high blood pressure? Data from a group of 6278 whites and blacks quizzed in the Third National Health and Nutritional Examination Survey between 1988 and 1994 (see CNN.com site article of January 6, 2000, titled “High Blood Pressure Greater Risk in U.S. South, Study Says”) indicated that a greater percentage of Southerners have high blood pressure than do people in any other region of the United States. This difference in rate of high blood pressure found in every ethnic group, gender, and age category studied. List at least two possible reasons we cannot conclude that living in the South causes high blood pressure.
There are many reasons for selecting a sample rather than obtaining information from an entire population (a census). Sometimes the process of measuring the characteristics of interest is destructive, as with measuring the lifetime of flashlight batteries or the sugar content of oranges, and it would be foolish to study the entire population. But the most common reason for selecting a sample is limited resources. Restrictions on available time or money usually prohibit observation of an entire population.

**Bias in Sampling**

Bias in sampling is the tendency for samples to differ from the corresponding population in some systematic way. Bias can result from the way in which the sample is selected or from the way in which information is obtained once the sample has been chosen. The most common types of bias encountered in sampling situations are selection bias, measurement or response bias, and nonresponse bias.

**Selection bias** (sometimes also called undercoverage) is introduced when the way the sample is selected systematically excludes some part of the population of interest. For example, a researcher may wish to generalize from the results of a study to the population consisting of all residents of a particular city, but the method of selecting individuals may exclude the homeless or those without telephones. If those who are excluded from the sampling process differ in some systematic way from those who are included, the sample is virtually guaranteed to be unrepresentative of the population. If this difference between the included and the excluded occurs on a variable that is important to the study, conclusions based on the sample data may not be valid for the population of interest. Selection bias also occurs if only volunteers or self-selected individuals are used in a study, because those who choose to participate (for example, in a call-in telephone poll) may well differ from those who choose not to participate.

**Measurement or response bias** occurs when the method of observation tends to produce values that systematically differ from the true value in some way. This might happen if an improperly calibrated scale is used to weigh items or if questions on a survey are worded in a way that tends to influence the response. For example, a Gallup survey sponsored by the American Paper Institute (Wall Street Journal, May 17, 1994) included the following question: “It is estimated that disposable diapers account for less than 2 percent of the trash in today’s landfills. In contrast, beverage containers, third-class mail and yard waste are estimated to account for about 21 percent of trash in landfills. Given this, in your opinion, would it be fair to tax or ban disposable diapers?” It is likely that the wording of this question prompted people to respond in a particular way.

Other things that might contribute to response bias are the appearance or behavior of the person asking the question, the group or organization conducting the study, and the tendency for people not to be completely honest when asked about illegal behavior or unpopular beliefs.

Although the terms measurement bias and response bias are often used interchangeably, the term measurement bias is usually used to describe systematic deviation from the true value as a result of a faulty measurement instrument (as with the improperly calibrated scale).

**Nonresponse bias** occurs when responses are not obtained from all individuals selected for inclusion in the sample. As with selection bias, nonresponse bias can distort
results if those who respond differ in important ways from those who do not respond. Although some level of nonresponse is unavoidable in most surveys, the biasing effect on the resulting sample is lowest when the response rate is high. To minimize nonresponse bias, it is critical that a serious effort be made to follow up with individuals who do not respond to an initial request for information.

The nonresponse rate for surveys or opinion polls varies dramatically, depending on how the data are collected. Surveys are commonly conducted by mail, by phone, and by personal interview. Mail surveys are inexpensive but often have high nonresponse rates. Telephone surveys can also be inexpensive and can be implemented quickly, but they work well only for short surveys and they can also have high nonresponse rates. Personal interviews are generally expensive but tend to have better response rates. Some of the many challenges of conducting surveys are discussed in Section 2.5.

### Types of Bias

| Selection Bias | Tendency for samples to differ from the corresponding population as a result of systematic exclusion of some part of the population. |
| Measurement or Response Bias | Tendency for samples to differ from the corresponding population because the method of observation tends to produce values that differ from the true value. |
| Nonresponse Bias | Tendency for samples to differ from the corresponding population because data are not obtained from all individuals selected for inclusion in the sample. |

It is important to note that bias is introduced by the way in which a sample is selected or by the way in which the data are collected from the sample. Increasing the size of the sample, though possibly desirable for other reasons, does nothing to reduce bias if the method of selecting the sample is flawed or if the nonresponse rate remains high. A good discussion of types of bias appears in the sampling book by Lohr listed in the references in the back of the book.

Potential sources of bias are illustrated in the following examples.

**EXAMPLE 2.1 Are Cell Phone Users Different?**

Many surveys are conducted by telephone and participants are often selected from phone books that include only landline telephones. For many years, it was thought that this was not a serious problem because most cell phone users also had a landline phone and so they still had a chance of being included in the survey. But the number of people with only cell phones is growing, and this trend is a concern for survey organizations. The article “Omitting Cell Phone Users May Affect Polls” (Associated Press, September 25, 2008) described a study that examined whether people who only have a cell phone are different than those who have landline phones. One finding from the study was that for people under the age of 30 with only a cell phone, 28% were Republicans compared to 36% of landline users. This suggests that researchers who use telephone surveys need to worry about how selection bias might influence the ability to generalize the results of a survey if only landlines are used.
The article “What People Buy from Fast-Food Restaurants: Caloric Content and Menu Item Selection” (Obesity [2009]: 1369–1374) reported that the average number of calories consumed at lunch in New York City fast food restaurants was 827. The researchers selected 267 fast food locations at random. The paper states that at each of these locations “adult customers were approached as they entered the restaurant and asked to provide their food receipt when exiting and to complete a brief survey.” Approaching customers as they entered the restaurant and before they ordered may have influenced what they purchased. This introduces the potential for response bias. In addition, some people chose not to participate when approached. If those who chose not to participate differed from those who did participate, the researchers also need to be concerned about nonresponse bias. Both of these potential sources of bias limit the researchers’ ability to generalize conclusions based on data from this study.

Random Sampling

Most of the inferential methods introduced in this text are based on the idea of random selection. The most straightforward sampling method is called simple random sampling. A simple random sample is a sample chosen using a method that ensures that each different possible sample of the desired size has an equal chance of being the one chosen. For example, suppose that we want a simple random sample of 10 employees chosen from all those who work at a large design firm. For the sample to be a simple random sample, each of the many different subsets of 10 employees must be equally likely to be the one selected. A sample taken from only full-time employees would not be a simple random sample of all employees, because someone who works part-time has no chance of being selected. Although a simple random sample may, by chance, include only full-time employees, it must be selected in such a way that each possible sample, and therefore every employee, has the same chance of inclusion in the sample. It is the selection process, not the final sample, which determines whether the sample is a simple random sample.

The letter \(n\) is used to denote sample size; it is the number of individuals or objects in the sample. For the design firm scenario of the previous paragraph, \(n = 10\).

**Definition**

A simple random sample of size \(n\) is a sample that is selected from a population in a way that ensures that every different possible sample of the desired size has the same chance of being selected.

The definition of a simple random sample implies that every individual member of the population has an equal chance of being selected. **However, the fact that every individual has an equal chance of selection, by itself, is not enough to guarantee that the sample is a simple random sample.** For example, suppose that a class is made up of 100 students, 60 of whom are female. A researcher decides to select 6 of the female students by writing all 60 names on slips of paper, mixing the slips, and then picking 6. She then selects 4 male students from the class using a similar procedure. Even though every student in the class has an equal chance of being included in the sample (6 of 60 females...
are selected and 4 of 40 males are chosen), the resulting sample is not a simple random sample because not all different possible samples of 10 students from the class have the same chance of selection. Many possible samples of 10 students—for example, a sample of 7 females and 3 males or a sample of all females—have no chance of being selected. The sample selection method described here is not necessarily a bad choice (in fact, it is an example of stratified sampling, to be discussed in more detail shortly), but it does not produce a simple random sample, and this must be considered when a method is chosen for analyzing data resulting from such a sampling method.

**Selecting a Simple Random Sample**

A number of different methods can be used to select a simple random sample. One way is to put the name or number of each member of the population on different but identical slips of paper. The process of thoroughly mixing the slips and then selecting \( n \) slips one by one yields a random sample of size \( n \). This method is easy to understand, but it has obvious drawbacks. The mixing must be adequate, and producing the necessary slips of paper can be extremely tedious, even for relatively small populations.

A commonly used method for selecting a random sample is to first create a list, called a **sampling frame**, of the objects or individuals in the population. Each item on the list can then be identified by a number, and a table of random digits or a random number generator can be used to select the sample. A random number generator procedure that produces a sequence of numbers that satisfies properties associated with the notion of randomness. Most statistics software packages include a random number generator, as do many calculators. A small table of random digits can be found in Appendix A, Table 1.

For example, suppose a list containing the names of the 427 customers who purchased a new car during 2009 at a large dealership is available. The owner of the dealership wants to interview a sample of these customers to learn about customer satisfaction. She plans to select a simple random sample of 20 customers. Because it would be tedious to write all 427 names on slips of paper, random numbers can be used to select the sample. To do this, we can use three-digit numbers, starting \( 001 \) and ending with \( 427 \), to represent the individuals on the list.

The random digits from rows 6 and 7 of Appendix A, Table 1 are shown here:

\[
0 \quad 9 \quad 3 \quad 8 \quad 7 \quad 6 \quad 7 \quad 9 \quad 9 \quad 5 \quad 6 \quad 2 \quad 5 \quad 6 \quad 5 \quad 8 \quad 4 \quad 2 \quad 6 \quad 4
\]
\[
4 \quad 1 \quad 0 \quad 1 \quad 0 \quad 2 \quad 2 \quad 0 \quad 4 \quad 7 \quad 5 \quad 1 \quad 1 \quad 9 \quad 4 \quad 7 \quad 9 \quad 7 \quad 5 \quad 1
\]

We can use blocks of three digits from this list (underlined in the lists above) to identify the individuals who should be included in the sample. The first block of three digits is 093, so the 93rd person on the list will be included in the sample. The next five blocks of three digits (876, 799, 562, 565, and 842) do not correspond to anyone on the list, so we ignore them. The next block that corresponds to a person on the list is 410, so that person is included in the sample. This process would continue until 20 people have been selected for the sample. We would ignore any three-digit repeats since any particular person should only be selected once for the sample.

Another way to select the sample would be to use computer software or a graphing calculator to generate 20 random numbers. For example, Minitab produced the following when 20 random numbers between 1 and 427 were requested.

\[
289 \quad 67 \quad 29 \quad 26 \quad 205 \quad 214 \quad 422 \quad 31 \quad 233 \quad 98
\]
\[
10 \quad 203 \quad 346 \quad 186 \quad 232 \quad 410 \quad 43 \quad 293 \quad 25 \quad 371
\]

These numbers could be used to determine which 20 customers to include in the sample.
When selecting a random sample, researchers can choose to do the sampling with or without replacement. **Sampling with replacement** means that after each successive item is selected for the sample, the item is “replaced” back into the population and may therefore be selected again at a later stage. In practice, sampling with replacement is rarely used. Instead, the more common method is to not allow the same item to be included in the sample more than once. After being included in the sample, an individual or object would not be considered for further selection. Sampling in this manner is called **sampling without replacement**.

**DEFINITION**

**Sampling without replacement**: Once an individual from the population is selected for inclusion in the sample, it may not be selected again in the sampling process. A sample selected without replacement includes \( n \) distinct individuals from the population.

**Sampling with replacement**: After an individual from the population is selected for inclusion in the sample and the corresponding data are recorded, the individual is placed back in the population and can be selected again in the sampling process. A sample selected with replacement might include any particular individual from the population more than once.

Although these two forms of sampling are different, when the sample size \( n \) is small relative to the population size, as is often the case, there is little practical difference between them. In practice, the two can be viewed as equivalent if the sample size is at most 10% of the population size.

**EXAMPLE 2.3 Selecting a Random Sample of Glass Soda Bottles**

Breaking strength is an important characteristic of glass soda bottles. Suppose that we want to measure the breaking strength of each bottle in a random sample of size \( n = 3 \) selected from four crates containing a total of 100 bottles (the population). Each crate contains five rows of five bottles each. We can identify each bottle with a number from 1 to 100 by numbering across the rows in each crate, starting with the top row of crate 1, as pictured:

<table>
<thead>
<tr>
<th>Crate 1</th>
<th>Crate 2</th>
<th>Crate 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
<td>26 27 28 ...</td>
<td>76 77 ...</td>
</tr>
<tr>
<td>6 ...</td>
<td>...</td>
<td>100</td>
</tr>
</tbody>
</table>

Using a random number generator from a calculator or statistical software package, we could generate three random numbers between 1 and 100 to determine which bottles would be included in our sample. This might result in bottles 15 (row 3 column 5 of crate 1), 89 (row 3 column 4 of crate 4), and 60 (row 2 column 5 of crate 3) being selected.
The goal of random sampling is to produce a sample that is likely to be representative of the population. Although random sampling does not guarantee that the sample will be representative, it does allow us to assess the risk of an unrepresentative sample. It is the ability to quantify this risk that will enable us to generalize with confidence from a random sample to the corresponding population.

An Important Note Concerning Sample Size

It is a common misconception that if the size of a sample is relatively small compared to the population size, the sample cannot possibly accurately reflect the population. Critics of polls often make statements such as, “There are 14.6 million registered voters in California. How can a sample of 1000 registered voters possibly reflect public opinion when only about 1 in every 14,000 people is included in the sample?” These critics do not understand the power of random selection!

Consider a population consisting of 5000 applicants to a state university, and suppose that we are interested in math SAT scores for this population. A dotplot of values in this population is shown in Figure 2.1(a). Figure 2.1(b) shows dotplots of math SAT scores for individuals in five different random samples from the population, ranging in sample size from \( n = 50 \) to \( n = 1000 \). Notice that the samples tend to reflect the distribution of scores in the population. If we were interested in using
sample to estimate the population average or to say something about the variability in SAT scores, even the smallest of the samples \( n = 50 \) pictured would provide reliable information. Although it is possible to obtain a simple random sample that does not do a reasonable job of representing the population, this is likely only when the sample size is very small, and unless the population itself is small, this risk does not depend on what fraction of the population is sampled. The random selection process allows us to be confident that the resulting sample adequately reflects the population, even when the sample consists of only a small fraction of the population.

**Other Sampling Methods**

Simple random sampling provides researchers with a sampling method that is objective and free of selection bias. In some settings, however, alternative sampling methods may be less costly, easier to implement, and sometimes even more accurate.

**Stratified Random Sampling** When the entire population can be divided into a set of nonoverlapping subgroups, a method known as stratified sampling often proves easier to implement and more cost-effective than simple random sampling. In stratified random sampling, separate simple random samples are independently selected from each subgroup. For example, to estimate the average cost of malpractice insurance, a researcher might find it convenient to view the population of all doctors practicing in a particular metropolitan area as being made up of four subpopulations: (1) surgeons, (2) internists and family practitioners, (3) obstetricians, and (4) a group that includes all other areas of specialization. Rather than taking a random simple sample from the population of all doctors, the researcher could take four separate simple random samples—one from the group of surgeons, another from the internists and family practitioners, and so on. These four samples would provide information about the four subgroups as well as information about the overall population of doctors.

When the population is divided in this way, the subgroups are called strata and each individual subgroup is called a stratum (the singular of strata). Stratified sampling entails selecting a separate simple random sample from each stratum. Stratified sampling can be used instead of simple random sampling if it is important to obtain information about characteristics of the individual strata as well as of the entire population, although a stratified sample is not required to do this—subgroup estimates can also be obtained by using an appropriate subset of data from a simple random sample.

The real advantage of stratified sampling is that it often allows us to make more accurate inferences about a population than does simple random sampling. In general, it is much easier to produce relatively accurate estimates of characteristics of a homogeneous group than of a heterogeneous group. For example, even with a small sample, it is possible to obtain an accurate estimate of the average grade point average (GPA) of students graduating with high honors from a university. The individual GPAs of these students are all quite similar (a homogeneous group), and even a sample of three or four individuals from this subpopulation should be representative. On the other hand, producing a reasonably accurate estimate of the average GPA of all seniors at the university, a much more diverse group of GPAs, is a more difficult task. Thus, if a varied population can be divided into strata, with each stratum being much more homogeneous than the population with respect to the characteristic of interest, then a stratified random sample can produce more accurate estimates of population characteristics than a simple random sample of the same size.

**Cluster Sampling** Sometimes it is easier to select groups of individuals from a population than it is to select individuals themselves. Cluster sampling involves dividing the population of interest into nonoverlapping subgroups, called clusters. Clusters
are then selected at random, and then all individuals in the selected clusters are included in the sample. For example, suppose that a large urban high school has 600 senior students, all of whom are enrolled in a first period homeroom. There are 24 senior homerooms, each with approximately 25 students. If school administrators wanted to select a sample of roughly 75 seniors to participate in an evaluation of the college and career placement advising available to students, they might find it much easier to select three of the senior homerooms at random and then include all the students in the selected homerooms in the sample. In this way, an evaluation survey could be administered to students in the selected homerooms at the same time—certainly easier logistically than randomly selecting 75 students and then administering the survey to those individual seniors.

Because whole clusters are selected, the ideal situation for cluster sampling is when each cluster mirrors the characteristics of the population. When this is the case, a small number of clusters results in a sample that is representative of the population. If not reasonable to think that the variability present in the population is reflected in the cluster, as is often the case when the cluster sizes are small, then it becomes important to ensure that a large number of clusters are included in the sample.

Be careful not to confuse clustering and stratification. Even though both of these sampling strategies involve dividing the population into subgroups, both the way in which the subgroups are sampled and the optimal strategy for creating the subgroups are different. In stratified sampling, we sample from every stratum, whereas in cluster sampling, we include only selected whole clusters in the sample. Because of this difference, to increase the chance of obtaining a sample that is representative of the population, we want to create homogeneous groups for strata and heterogeneous (reflecting the variability in the population) group clusters.

**Systematic Sampling**

**Systematic sampling** is a procedure that can be used when it is possible to view the population of interest as consisting of a list or some other sequential arrangement. A value \( k \) is specified (for example, \( k = \frac{50}{200} \) or \( k = \frac{200}{200} \)). The first of the first \( k \) individuals is selected at random, after which every \( k \)th individual in the sequence is included in the sample. A sample selected in this way is called a **1 in \( k \) systematic sample**.

For example, a sample of faculty members at a university might be selected from the faculty phone directory. One of the first \( k \) faculty members listed could be selected at random, and then every 20th faculty member after that on the list would also be included in the sample. This would result in a 1 in 20 systematic sample.

The value of \( k \) for a 1 in \( k \) systematic sample is generally chosen to achieve desired sample size. For example, in the faculty directory scenario just described, if there were 900 faculty members at the university, the 1 in 20 systematic sample described would result in a sample size of 45. If a sample size of 100 was desired, a 1 in 9 systematic sample could be used (because \( 900/100 = 9 \)).

As long as there are no repeating patterns in the population list, systematic sampling works reasonably well. However, if there are such patterns, systematic sampling can result in an unrepresentative sample. For example, suppose that workers at an entry station of a state park have recorded the number of visitors to the park each day for the past 10 years. In a 1 in 70 systematic sample of days from this list, we pick one of the first 70 days at random and then every 70th day after that. But if first day selected happened to be a Wednesday, every day selected in the entire sample would also be a Wednesday (because there are 7 days a week and 70 is a multiple 7). It is unlikely that such a sample would be representative of the entire collection of days. The number of visitors is likely to be higher on weekend days, and no Saturdays or Sundays would be included in the sample.
Convenience Sampling: Don’t Go There! It is often tempting to resort to “convenience” sampling—that is, using an easily available or convenient group to form a sample. This is a recipe for disaster! Results from such samples are rarely informative, and it is a mistake to try to generalize from a convenience sample to any larger population.

One common form of convenience sampling is sometimes called voluntary response sampling. Such samples rely entirely on individuals who volunteer to be a part of the sample, often by responding to an advertisement, calling a publicized telephone number to register an opinion, or logging on to an Internet site to complete a survey. It is extremely unlikely that individuals participating in such voluntary response surveys are representative of any larger population of interest.

**EXERCISES 2.13 - 2.32**

2.13 As part of a curriculum review, the psychology department would like to select a simple random sample of 20 of last year’s 140 graduates to obtain information on how graduates perceived the value of the curriculum. Describe two different methods that might be used to select the sample.

2.14 A petition with 500 signatures is submitted to a university’s student council. The council president would like to determine the proportion of those who signed the petition who are actually registered students at the university. There is not enough time to check all 500 names with the registrar, so the council president decides to select a simple random sample of 30 signatures. Describe how this might be done.

2.15 During the previous calendar year, a county’s small claims court processed 870 cases. Describe how a simple random sample of size $n = 50$ might be selected from the case files to obtain information regarding the average award in such cases.

2.16 The financial aid advisor of a university plans to use a stratified random sample to estimate the average amount of money that students spend on textbooks each term. For each of the following proposed stratification schemes, discuss whether it would be worthwhile to stratify the university students in this manner.
   a. Strata corresponding to class standing (freshman, sophomore, junior, senior, graduate student)
   b. Strata corresponding to field of study, using the following categories: engineering, architecture, business, other
   c. Strata corresponding to the first letter of the last name: A–E, F–K, etc.

2.17 Suppose that a group of 1000 orange trees is laid out in 40 rows of 25 trees each. To determine the sugar content of fruit from a sample of 30 trees, researcher A suggests randomly selecting five rows and then randomly selecting six trees from each sampled row. Researcher B suggests numbering each tree on a map of the trees from 1 to 1000 and using random numbers to select 30 of the trees. Which selection method is preferred? Explain.

2.18 For each of the situations described, state whether the sampling procedure is simple random sampling, stratified random sampling, cluster sampling, systematic sampling, or convenience sampling.
   a. All first-year students at a university are enrolled in 1 of 30 sections of a seminar course. To select a sample of freshmen at this university, a researcher selects four sections of the seminar course at random from the 30 sections and all students in the four selected sections are included in the sample.
   b. To obtain a sample of students, faculty, and staff at a university, a researcher randomly selects 50 faculty members from a list of faculty, 100 students from a list of students, and 30 staff members from a list of staff.
   c. A university researcher obtains a sample of students at his university by using the 85 students enrolled in his Psychology 101 class.
   d. To obtain a sample of the seniors at a particular high school, a researcher writes the name of each senior on a slip of paper, places the slips in a box and mixes them, and then selects 10 slips. The students whose names are on the selected slips of paper are included in the sample.
   e. To obtain a sample of those attending a basketball game, a researcher selects the 24th person through the door. Then, every 50th person after that is also included in the sample.

Bold exercises answered in back  ● Data set available online  ➤ Video Solution available
2.19 Of the 6500 students enrolled at a community college, 3000 are part time and the other 3500 are full time. The college can provide a list of students that is sorted so that all full-time students are listed first, followed by the part-time students.
   a. Describe a procedure for selecting a stratified random sample that uses full-time and part-time students as the two strata and that includes 10 students from each stratum.
   b. Does every student at this community college have the same chance of being selected for inclusion in the sample? Explain.

2.20 Briefly explain why it is advisable to avoid the use of convenience samples.

2.21 A sample of pages from this book is to be obtained, and the number of words on each selected page will be determined. For the purposes of this exercise, equations are not counted as words and a number is counted as a word only if it is spelled out—that is, ten is counted as a word, but 10 is not.
   a. Describe a sampling procedure that would result in a simple random sample of pages from this book.
   b. Describe a sampling procedure that would result in a stratified random sample. Explain why you chose the specific strata used in your sampling plan.
   c. Describe a sampling procedure that would result in a systematic sample.
   d. Describe a sampling procedure that would result in a cluster sample.
   e. Using the process you gave in Part (a), select a simple random sample of at least 20 pages, and record the number of words on each of the selected pages. Construct a dotplot of the resulting sample values, and write a sentence or two commenting on what it reveals about the number of words on a page.
   f. Using the process you gave in Part (b), select a stratified random sample that includes a total of at least 20 selected pages, and record the number of words on each of the selected pages. Construct a dotplot of the resulting sample values, and write a sentence or two commenting on what it reveals about the number of words on a page.

2.22 In 2000, the chairman of a California ballot initiative campaign to add “none of the above” to the list of ballot options in all candidate races was quite critical of a Field poll that showed his measure trailing by 10 percentage points. The poll was based on a random sample of 1000 registered voters in California. He is quoted by the Associated Press (January 30, 2000) as saying “Field’s sample in that poll equates to one out of 17, voters,” and he added that this was so dishonest Field should get out of the polling business! If worked on the Field poll, how would you respond to criticism?

2.23 The authors of the paper “Digital Inequality: Differences in Young Adults’ Use of the Internet” (Computer-Human Interaction [2009]: 130–136) described an investigation into whether lying is less common in face-to-face communication than in other forms of communication such as phone conversations or e-mail. Participants in this study were 30 students in an upper-division communications course at Cornell University who received course credit for participation. Participants were asked to record all of their social interactions for a week, making note of any lies told. Based on data from these records, the authors of the paper concluded students lie more often in phone conversations than in face-to-face conversations and more often in face-to-face conversations than in e-mail. Discuss the limitations of this study, commenting on the way the sample was selected and potential sources of bias.

2.24 The authors of the paper “Illicit Use of Psycho-stimulants among College Students” (Psychology, Health & Medicine [2002]: 283–287) surveyed college students about their use of legal and illegal stimulants. The sample of students surveyed consisted of students enrolled in psychology class at a small, competitive college in the United States.
   a. Was this sample a simple random sample, a stratified random sample, or a convenience sample? Explain.
   b. Give two reasons why the estimate of the proportion of students who reported using illegal stimulants based on data from this survey should not be generalized to all U.S. college students.

2.25 The paper “Deception and Design: The Impact of Communication Technology on Lying Behavior” (Computer-Human Interaction [2009]: 130–136) describes an investigation into whether lying is less common in face-to-face communication than in other forms of communication such as phone conversations or e-mail. Participants in this study were 30 students in an upper-division communications course at Cornell University who received course credit for participation. Participants were asked to record all of their social interactions for a week, making note of any lies told. Based on data from these records, the authors of the paper concluded students lie more often in phone conversations than in face-to-face conversations and more often in face-to-face conversations than in e-mail. Discuss the limitations of this study, commenting on the way the sample was selected and potential sources of bias.
2.26 The authors of the paper "Popular Video Games: Quantifying the Presentation of Violence and its Context" (Journal of Broadcasting & Electronic Media [2003]: 58–76) investigated the relationship between video game rating—suitable for everyone (E), suitable for 13 years of age and older (T), and suitable for 17 years of age and older (M)—and the number of violent interactions per minute of play. The sample of games examined consisted of 60 video games—the 20 most popular (by sales) for each of three game systems. The researchers concluded that video games rated for older children had significantly more violent interactions per minute than did those games rated for more general audiences.

a. Do you think that the sample of 60 games was selected in a way that makes it reasonable to think it is representative of the population of all video games?

b. Is it reasonable to generalize the researchers' conclusions to all video games? Explain why or why not.

2.27 Participants in a study of honesty in online dating profiles were recruited through print and online advertisements in the Village Voice, one of New York City’s most prominent weekly newspapers, and on Craigslist New York City ("The Truth About Lying in Online Dating Profiles," Computer–Human Interaction [2007]: 1–4). The actual height, weight, and age of the participants were compared to what appeared in their online dating profiles. The resulting data was then used to draw conclusions about how common deception was in online dating profiles. What concerns do you have about generalizing conclusions based on data from this study to the population of all people who have an online dating profile? Be sure to address at least two concerns and give the reason for your concern.

2.28 The report "Undergraduate Students and Credit Cards in 2004: An Analysis of Usage Rates and Trends" (Nellie Mae, May 2005) estimated that 21% of undergraduates with credit cards pay them off each month and that the average outstanding balance on undergraduates’ credit cards is $2169. These estimates were based on an online survey that was sent to 1260 students. Responses were received from 132 of these students. Is it reasonable to generalize the reported estimates to the population of all undergraduate students? Address at least two possible sources of bias in your answer.

2.29 Suppose that you were asked to help design a survey of adult city residents in order to estimate the proportion that would support a sales tax increase. The plan is to use a stratified random sample, and three stratification schemes have been proposed.

Scheme 1: Stratify adult residents into four strata based on the first letter of their last name (A–G, H–N, O–T, U–Z).

Scheme 2: Stratify adult residents into three strata: college students, nonstudents who work full time, nonstudents who do not work full time.

Scheme 3: Stratify adult residents into five strata by randomly assigning residents into one of the five strata.

Which of the three stratification schemes would be best in this situation? Explain.

2.30 The article "High Levels of Mercury Are Found in Californians" (Los Angeles Times, February 9, 2006) describes a study in which hair samples were tested for mercury. The hair samples were obtained from more than 6000 people who voluntarily sent hair samples to researchers at Greenpeace and The Sierra Club. The researchers found that nearly one-third of those tested had mercury levels that exceeded the concentration thought to be safe. Is it reasonable to generalize this result to the larger population of U.S. adults? Explain why or why not.

2.31 Whether or not to continue a Mardi Gras Parade through downtown San Luis Obispo, CA, is a hotly debated topic. The parade is popular with students and many residents, but some celebrations have led to complaints and a call to eliminate the parade. The local newspaper conducted online and telephone surveys of its readers and was surprised by the results. The survey web site received more than 400 responses, with more than 60% favoring continuing the parade, while the telephone response line received more than 120 calls, with more than 90% favoring banning the parade (San Luis Obispo Tribune, March 3, 2004). What factors may have contributed to these very different results?

2.32 The article "Gene’s Role in Cancer May Be Overstated" (San Luis Obispo Tribune, August 21, 2002) states that “early studies that evaluated breast cancer risk among gene mutation carriers selected women in families where sisters, mothers, and grandmothers all had breast cancer. This created a statistical bias that skewed risk estimates for women in the general population.” Is the bias described here selection bias, measurement bias, or nonresponse bias? Explain.
Sometimes the questions we are trying to answer deal with the effect of certain explanatory variables on some response. Such questions are often of the form, “What happens when . . . ?” or “What is the effect of . . . ?” For example, an industrial engineer may be considering two different workstation designs and might want to know whether the choice of design affects work performance. A medical researcher may want to determine how a proposed treatment for a disease compares to a standard treatment. Experiments provide a way to collect data to answer these type questions.

**DEFINITION**

An **experiment** is a study in which one or more explanatory variables are manipulated in order to observe the effect on a response variable.

The **explanatory variables** are those variables that have values that are controlled by the experimenter. Explanatory variables are also called **factors**.

The **response variable** is a variable that is not controlled by the experimenter and that is measured as part of the experiment.

An **experimental condition** is any particular combination of values for the explanatory variables. Experimental conditions are also called **treatments**.

Suppose we are interested in determining the effect of room temperature on performance on a first-year calculus exam. In this case, the explanatory variable is room temperature (it can be manipulated by the experimenter). The response variable is exam performance (the variable that is not controlled by the experimenter and will be measured).

In general, we can identify the explanatory variables and the response variable easily if we can describe the purpose of the experiment in the following terms:

The purpose is to assess the effect of \[ \text{explanatory variable} \] on \[ \text{response variable} \].

Let’s return to the example of an experiment to assess the effect of room temperature on exam performance. We might decide to use two room temperature settings, 65° and 75°. This would result in an experiment with two experimental conditions (or equivalently, two treatments) corresponding to the two temperature settings.

Suppose that there are 10 sections of first-semester calculus that have agreed to participate in our study. We might design an experiment in this way: Set the room temperature (in degrees Fahrenheit) to 65° in five of the rooms and to 75° in the other five rooms on test day, and then compare the exam scores for the 65° group and the 75° group. Suppose that the average exam score for the students in the 65° group was noticeably higher than the average for the 75° group. Could we conclude that the increased temperature resulted in a lower average score? Based on the information given, the answer is no because many other factors might be related to exam score. Were the sections at different times of the day? Did they have the same instructor? Different textbooks? Did the sections differ with respect to the abilities of the students? Any of these other factors could provide a plausible explanation (having nothing to do with room temperature) for why the average test score was different for the two groups. It is not possible to separate the effect of temperature from the effects of these other factors.
these other factors. As a consequence, simply setting the room temperatures as described makes for a poorly designed experiment.

A well-designed experiment requires more than just manipulating the explanatory variables; the design must also eliminate other possible explanations or the experimental results will not be conclusive.

The goal is to design an experiment that will allow us to determine the effects of the explanatory variables on the chosen response variable. To do this, we must take into consideration any extraneous variables that, although not of interest in the current study, might also affect the response variable.

**DEFINITION**

An **extraneous variable** is one that is not one of the explanatory variables in the study but is thought to affect the response variable.

A well-designed experiment copes with the potential effects of extraneous variables by using random assignment to experimental conditions and sometimes also by incorporating direct control and/or blocking into the design of the experiment. Each of these strategies—random assignment, direct control, and blocking—is described in the paragraphs that follow.

A researcher can **directly control** some extraneous variables. In the calculus test example, the textbook used is an extraneous variable because part of the differences in test results might be attributed to this variable. We could control this variable directly, by requiring that all sections use the same textbook. Then any observed differences between temperature groups could not be explained by the use of different textbooks. The extraneous variable **time of day** might also be directly controlled in this way by having all sections meet at the same time.

The effects of some extraneous variables can be filtered out by a process known as **blocking**. Extraneous variables that are addressed through blocking are called **blocking variables**. Blocking creates groups (called blocks) that are similar with respect to blocking variables; then all treatments are tried in each block. In our example, we might use **instructor** as a blocking variable. If five instructors are each teaching two sections of calculus, we would make sure that for each instructor, one section was part of the 65° group and the other section was part of the 75° group. With this design, if we see a difference in exam scores for the two temperature groups, the extraneous variable **instructor** can be ruled out as a possible explanation, because all five instructors’ students were present in each temperature group. (Had we controlled the instructor variable by choosing to have only one instructor, that would be an example of direct control. Of course we can’t directly control both time of day and instructor.) If one instructor taught all the 65° sections and another taught all the 75° sections, we would be unable to distinguish the effect of temperature from the effect of the instructor. In this situation, the two variables (temperature and instructor) are said to be **confounded**.

Two variables are **confounded** if their effects on the response variable cannot be distinguished from one another.
If an extraneous variable is confounded with the explanatory variables (define the treatments), it is not possible to draw an unambiguous conclusion at the effect of the treatment on the response. Both direct control and blocking are effective in ensuring that the controlled variables and blocking variables are not confounded with the variables that define the treatments.

We can directly control some extraneous variables by holding them constant, and we can use blocking to create groups that are similar to essentially filter out the effect of others. But what about variables, such as student ability in our calculus test example, which cannot be controlled by the experimenter and which would be difficult to use as blocking variables? These extraneous variables are handled by the use of random assignment to experimental groups. Random assignment ensures that our experimenter does not systematically favor one experimental condition over any other and attempt to create experimental groups that are as much alike as possible. For example, if students requesting calculus could be assigned to one of the ten available sections using a random mechanism, we would expect the resulting groups to be similar with respect to student ability as well as with respect to other extraneous variables that are not directly controlled or used as a basis for blocking. Note that random assignment in an experiment is different from random selection of subjects. The ideal situation would be to have both random selection of subjects and random assignment of subject experimental conditions, as this would allow conclusions from the experiment to be generalized to a larger population. For many experiments the random selection of subjects is not possible. As long as subjects are assigned at random to experimental conditions, it is still possible to assess treatment effects.

To get a sense of how random assignment tends to create similar groups, suppose that 50 college freshmen are available to participate as subjects in an experiment to investigate whether completing an online review of course material before an exam improves exam performance. The 50 subjects vary quite a bit with respect to achievement, which is reflected in their math and verbal SAT scores, as shown in Figure 2.2.

If these 50 students are to be assigned to the two experimental groups (one will complete the online review and one that will not), we want to make sure that assignment of students to groups does not favor one group over the other by tending to assign the higher achieving students to one group and the lower achieving students to the other.

Creating groups of students with similar achievement levels in a way that considers both verbal and math SAT scores simultaneously would be difficult, so we rely on random assignment. Figure 2.3(a) shows the math SAT scores of the students assigned to each of the two experimental groups (one shown in orange and one shown in blue) for each of three different random assignments of students to groups. Figure 2.3(b) shows the verbal SAT scores for the two experimental groups for each of the same three random assignments. Notice that each of the three random assignment produced groups that are similar with respect to both verbal and math SAT scores. If any of these three assignments were used and the two groups differed on exam performance, we could rule out differences in math or verbal SAT scores as potential competing explanations for the difference.
Collecting Data Sensibly

Not only will random assignment tend to create groups that are similar with respect to verbal and math SAT scores, but it will also tend to even out the groups with respect to other extraneous variables. As long as the number of subjects is not too small, we can rely on the random assignment to produce comparable experimental groups. This is the reason that random assignment is a part of all well-designed experiments.

Not all experiments require the use of human subjects. For example, a researcher interested in comparing three different gasoline additives with respect to gasoline mileage might conduct an experiment using a single car with an empty tank. One gallon of gas with one of the additives will be put in the tank, and the car will be driven along a standard route at a constant speed until it runs out of gas. The total distance traveled on the gallon of gas could then be recorded. This could be repeated a number of times—10, for example—with each additive.

The experiment just described can be viewed as consisting of a sequence of trials. Because a number of extraneous variables (such as variations in environmental conditions like wind speed or humidity and small variations in the condition of the car) might have an effect on gas mileage, it would not be a good idea to use additive 1 for the first 10 trials, additive 2 for the next 10 trials, and so on. A better approach would be to randomly assign additive 1 to 10 of the 30 planned trials, and then randomly assign additive 2 to 10 of the remaining 20 trials. The resulting plan for carrying out the experiment might look as follows:

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>...</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additive</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>...</td>
<td>1</td>
</tr>
</tbody>
</table>

When an experiment can be viewed as a sequence of trials, random assignment involves the random assignment of treatments to trials. Remember that random assignment—either of subjects to treatments or of treatments to trials—is a critical component of a good experiment.

Random assignment can be effective only if the number of subjects or observations in each experimental condition (treatment) is large enough for each experimental group to reliably reflect variability in the population. For example, if there were only 20 students requesting calculus, it is unlikely that we would get equivalent groups for comparison, even with random assignment to the ten sections. Replication is the design strategy of making multiple observations for each experimental condition. Together, replication and random assignment allow the researcher to be reasonably confident of comparable experimental groups.

To illustrate the design of a simple experiment, consider the dilemma of Anna, a waitress in a local restaurant. She would like to increase the amount of her tips, and her strategy is simple: She will write “Thank you” on the back of some of the checks before giving them to the patrons and on others she will write nothing. She plans to calculate the percentage of the tip as her measure of success (for example, a 15% tip is common). She will compare the average percentage of the tips calculated from...
2.3 Simple Comparative Experiments

checks with and without the handwritten “Thank you.” If writing “Thank you” does not produce higher tips, she may try a different strategy.

Anna is untrained in the art of planning experiments, but already she has taken some common sense steps in the right direction to answer her question—Will writing “Thank you” produce the desired outcome of higher tips? Anna has defined a manageable problem, and collecting the appropriate data is feasible. It should be easy to gather data as a normal part of her work. Anna wonders whether writing “Thank you” on customers’ bills will have an effect on the amount of her tip. In the language of experimentation, we would refer to the writing of “Thank you” and the not writing “Thank you” as treatments (the two experimental conditions to be compared in the experiment). The two treatments together are the possible values of the explanatory variable. The tipping percentage is the response variable. The idea behind this terminology is that the tipping percentage is a response to the treatments writing “Thank you” or not writing “Thank you.” Anna’s experiment may be thought of as an attempt to explain the variability in the response variable in terms of its presumed cause, the variability in the explanatory variable. That is, as she manipulates the explanatory variable, she expects the response by her customers to vary. Anna has a good start, but now she must consider the four fundamental design principles.

**Replication.** Anna cannot run a successful experiment by gathering tipping information on only one person for each treatment. There is no reason to believe that any single tipping incident is representative of what would happen in other incidents, and therefore it would be impossible to evaluate the two treatments with only two subjects. To interpret the effects of a particular treatment, she replicates each treatment in the experiment.

**Blocking.** Suppose that Anna works on both Thursdays and Fridays. Because day of the week might affect tipping behavior, Anna should block on day of week and make sure that observations for both treatments are made on each of the two days.
Direct Control and Random Assignment. There are a number of extraneous variables that might have an effect on the size of tip. Some restaurant patrons will be seated near the window with a nice view; some will have to wait for a table, whereas others may be seated immediately; and some may be on a fixed income and cannot afford a large tip. Some of these variables can be directly controlled. For example, Anna may choose to use only window tables in her experiment, thus eliminating table location as a potential confounding variable. Other variables, such as length of wait and customer income, cannot be easily controlled. As a result, it is important that Anna use random assignment to decide which of the window tables will be in the “Thank you” group and which will be in the “No thank you” group. She might do this by flipping a coin as she prepares the check for each window table. If the coin lands with the head side up, she could write “Thank you” on the bill, omitting the “Thank you” when a tail is observed.

The accompanying box summarizes how experimental designs deal with extraneous variables.

### Taking Extraneous Variables into Account

Extraneous variables are variables other than the explanatory variables in an experiment that may also have an effect on the response variable. There are several strategies for dealing with extraneous variables in order to avoid confounding.

Extraneous variables that we know about and choose to incorporate into the experimental design:

**Strategies**
- Direct control—holds extraneous variables fixed so that they can’t affect the response variable
- Blocking—allows for valid comparisons because each treatment is tried in each block

Extraneous variables that we don’t know about or choose not to incorporate into the experimental design through direct control or blocking:

**Strategy**
- Random assignment

Extraneous variables that are not incorporated into the design of the experiment are sometimes called lurking variables.*


---

A Note on Random Assignment

There are several strategies that can be used to perform random assignment of subjects to treatments or treatments to trials. Two common strategies are:

- Write the name of each subject or a unique number that corresponds to a subject on a slip of paper. Place all of the slips in a container and mix well. Then draw out the desired number of slips to determine those that will be assigned to the first treatment group. This process of drawing slips of paper then continues until all treatment groups have been determined.
• Assign each subject a unique number from 1 to \( n \), where \( n \) represents the total number of subjects. Use a random number generator or table of random numbers to obtain numbers that will identify which subjects will be assigned to the first treatment group. This process would be repeated, ignoring any random numbers generated that correspond to subjects that have already been assigned a treatment group, until all treatment groups have been formed.

The two strategies above work well and can be used for experiments in which desired number of subjects in each treatment group has been predetermined.

Another strategy that is sometimes employed is to use a random mechanism (such as tossing a coin or rolling a die) to determine which treatment will be assigned to a particular subject. For example, in an experiment with two treatments, you might toss a coin to determine if the first subject is assigned to treatment 1 or treatment 2. This could continue for each subject—if the coin lands H, the subject is assigned treatment 1, and if the coin lands T, the subject is assigned to treatment 2. This strategy is fine, but may result in treatment groups of unequal size. For example, in an experiment with 100 subjects, 53 might be assigned to treatment 1 and 47 to treatment 2. If this is acceptable, the coin flip strategy is a reasonable way to assign subjects to treatments.

But, suppose you want to ensure that there is an equal number of subjects in each treatment group. Is it acceptable to use the coin flip strategy until one treatment group is complete and then just assign all of the remaining subjects to groups that are not yet full? The answer to this question is that it is probably not acceptable. For example, suppose a list of 20 subjects is in order by age from youngest to oldest and that we want to form two treatment groups each consisting of 10 subjects. Toss a coin to make the assignments might result in the following (based on using the row of random digits in Appendix A, Table 1, with an even number representing H and an odd number representing T):

<table>
<thead>
<tr>
<th>Subject</th>
<th>Random Number</th>
<th>Coin Toss Equivalent</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>T</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td>Treatment group 2 filled. Assign all others to treatment group 1.</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
If the list of subjects was ordered by age, treatment group 1 would end up with a disproportionate number of older people. This strategy usually results in one treatment group drawing disproportionately from the end of the list. So, the only time the strategy of assigning at random until groups fill up and then assigning the remaining subjects to the group that is not full is reasonable is if you can be sure that the list is in random order with respect to all variables that might be related to the response variable. Because of this, it is best to avoid this strategy. Activity 2.5 investigates potential difficulties with this type of strategy.

On the other hand, if the number of subjects is large, it may not be important that every treatment group has exactly the same number of subjects. If this is the case, it is reasonable to use a coin flip strategy (or other strategies of this type) that does not involve stopping assignment of subjects to a group that becomes full.

**Evaluating an Experimental Design**

The key concepts of experimental design provide a framework for evaluating an experimental design, as illustrated in the following examples.

**EXAMPLE 2.4 Revenge Is Sweet**

The article “The Neural Basis of Altruistic Punishment” (Science, August 27, 2004) described a study that examined motivation for revenge. Subjects in the study were all healthy, right-handed men. Subjects played a game with another player in which they could both earn money by trusting each other or one player could double-cross the other player and keep all of the money. In some cases the double cross was required by the rules of the game in certain circumstances, while in other cases the double cross was the result of a deliberate choice. The victim of a double cross was then given the opportunity to retaliate by imposing a fine, but sometimes the victim had to spend some of his own money in order to impose the fine. This study was an experiment with four experimental conditions or treatments:

1. double cross not deliberate (double cross dictated by the rules of the game) and no cost to the victim to retaliate
2. double cross deliberate and no cost to the victim to retaliate
3. double cross not deliberate and a cost to the victim to retaliate
4. double cross deliberate and a cost to the victim to retaliate

All subjects chose revenge (imposed a fine on the double-crosser) when the double cross was deliberate and retaliation was free, and 86% of the subjects chose revenge when the double cross was deliberate, even if it cost them money. Only 21% imposed a fine if the double cross was dictated by the rules of the game and was not deliberate.

Assuming that the researchers randomly assigned the subjects to the four experimental conditions, this study is an experiment that incorporated random assignment, direct control (controlled sex, health, and handedness by using only healthy, right-handed males as subjects), and replication (many subjects assigned to each experimental condition).
EXAMPLE 2.5 Subliminal Messages

The article “The Most Powerful Manipulative Messages Are Hiding in Plain Sight (Chronicle of Higher Education, January 29, 1999) reported the results of an interesting experiment on priming—the effect of subliminal messages on how we behave.

In the experiment, subjects completed a language test in which they were asked to construct a sentence using each word in a list of words. One group of subjects received a list of words related to politeness, and a second group was given a list of words related to rudeness. Subjects were told to complete the language test and then come into the hall and find the researcher so that he could explain the next part of the test. When each subject came into the hall, he or she found the researcher engaged in conversation. The researcher wanted to see whether the subject would interrupt the conversation. The researcher found that 63% of those primed with words related to rudeness interrupted the conversation, whereas only 17% of those primed with words related to politeness interrupted.

If we assume that the researcher randomly assigned the subjects to the groups, then this study is an experiment that compares two treatments (primed with words related to rudeness and primed with words related to politeness). The response variable, politeness, has the values interrupted conversation and did not interrupt conversation. The experiment uses replication (many subjects in each treatment group) and random assignment to control for extraneous variables that might affect the response.

Many experiments compare a group that receives a particular treatment to a control group that receives no treatment.

EXAMPLE 2.6 Chilling Newborns? Then You Need a Control Group...

Researchers for the National Institute of Child Health and Human Development studied 208 infants whose brains were temporarily deprived of oxygen as a result of complications at birth (The New England Journal of Medicine, October 2005). These babies were subjects in an experiment to determine if reducing body temperature for three days after birth improved their chances of surviving without brain damage. The experiment was summarized in a paper that stated “infants were randomly assigned to usual care (control group) or whole-body cooling. Including a control group in the experiment provided a basis for comparison of death and disability rates for the proposed cooling treatment and those for usual care. Some extraneous variables that might also affect death and disability, such as the duration of oxygen deprivation, could not be directly controlled, so it is important that the experiment did not unintentionally favor one experimental condition over the other, random assignment of the infants to the two groups was critical. Because this was a well-designed experiment, the researchers were able to conclude that cooling did reduce the risk of death and disability for infants deprived of oxygen at birth.
Visualizing the Underlying Structure of Some Common Experimental Designs

Simple diagrams are sometimes used to highlight important features of some common experimental designs. The structure of an experiment that is based on random assignment of experimental units (the units to which treatments are assigned, usually subjects or trials) to one of two treatments is displayed in Figure 2.4. The diagram can be easily adapted for an experiment with more than two treatments. In any particular setting, we would also want to customize the diagram by indicating what the treatments are and what response will be measured. This is illustrated in Example 2.7.

Can moving their hands help children learn math? This is the question investigated by the authors of the paper "Gesturing Gives Children New Ideas about Math" (Psychological Science [2009]: 267–272). An experiment was conducted to compare two different methods for teaching children how to solve math problems of the form $3 + 2 + 8 = \_ \_ + 8$. One method involved having students point to the $3 + 2$ on the left side of the equal sign with one hand and then point to the blank on the right side of the equal sign before filling in the blank to complete the equation. The other method did not involve using these hand gestures. The paper states that the study used children ages 9 and 10 who were given a pretest containing six problems of the type described above. Only children who answered all six questions incorrectly became subjects in the experiment. There were a total of 128 subjects.

To compare the two methods, the 128 children were assigned at random to the two experimental conditions. Children assigned to one experimental condition were taught a method that used hand gestures and children assigned to the other experimental condition were taught a similar strategy that did not involve using hand gestures. Each child then took a test with six problems and the number correct was determined for each child. The researchers used the resulting data to reach the conclusion that the average number correct for children who used the method that incorporated hand gestures was significantly higher than the average number correct for children who were taught the method that did not use hand gestures.
The basic structure of this experiment can be diagramed as shown in Figure 2.5. This type of diagram provides a nice summary of the experiment, but notice several important characteristics of the experiment are not captured in the diagram. For example, the diagram does not show that some extraneous variables were considered by the researchers and directly controlled. In this example, both age and prior math knowledge were directly controlled by using only children who were 9 to 10 years old and who were not able to solve any of the questions on the pretest correctly. So, be aware that while a diagram of an experiment may be a useful tool, it cannot stand alone in describing an experimental design.

Some experiments consist of a sequence of trials, and treatments are assigned at random to the trials. The diagram in Figure 2.6 illustrates the underlying structure of such an experiment. Example 2.8 shows how this diagram can be customized to describe a particular experiment.
The paper “Effect of Cell Phone Distraction on Pediatric Pedestrian Injury Risk” (*Pediatrics* 2009: e179–e185) describes an experiment to investigate whether pedestrians who are talking on a cell phone are at greater risk of an accident when crossing the street than when not talking on a cell phone. No children were harmed in this experiment—a virtual interactive pedestrian environment was used! One possible way of conducting such an experiment would be to have a person cross 20 streets in this virtual environment. The person would talk on a cell phone for some crossings and would not use the cell phone for others. It would be important to randomly assign the two treatments (talking on the phone, not talking on the phone) to the 20 trials (the 20 simulated street crossings). This would result in a design that did not favor one treatment over the other because the pedestrian became more careful with experience or more tired and, therefore, easily distracted over time. The basic structure of this experiment is diagramed in Figure 2.7.

The actual experiment conducted by the authors of the paper was a bit more sophisticated than the one just described. In this experiment, 77 children age 10 and 11 each performed simulated crossings with and without a cell phone. Random assignment was used to decide which children would cross first with the cell phone followed by no cell phone and which children could cross first with no cell phone. The structure of this experiment is diagramed in Figure 2.8.
As was the case in Example 2.7, note that while the diagram is informative itself, it does not capture all of the important aspects of the design. In particular, it does not capture the direct control of age (only children age 10 and 11 were used as subjects in the experiment).

Experimental designs in which experimental units are assigned at random to treatments or in which treatments are assigned at random to trials (like those of experiments in Examples 2.7 and 2.8) are called **completely randomized design**.

Diagrams are also useful for highlighting the structure of experiments that use blocking. This is illustrated in Example 2.9.

---

**EXAMPLE 2.9 A Helping Hand Revisited**

Let’s return to the experiment described in Example 2.7. Take a minute to go back and re-read that example. The experiment described in Example 2.7, a completely randomized design with 128 subjects, was used to compare two different methods for teaching kids how to solve a particular type of math problem. Age and prior math knowledge were extraneous variables that the researchers thought might be related to performance on the math test given at the end of the lesson, so the researchers chose to directly control these variables. The 128 children were assigned at random to the two experimental conditions (treatments). The researchers relied on random assignment to create treatment groups that would be roughly equivalent with respect to other extraneous variables.

But suppose that we were worried that gender might also be related to performance on the math test. One possibility would be to use direct control of gender—that is, we might use only boys or only girls as subjects in the experiment. Then if we saw a difference in test performance for the two teaching methods, it could not be due to one experimental group containing more boys and fewer girls than the other group. The downside to this strategy is that if we use only boys in the experiment, there is no basis for also generalizing any conclusions from the experiment to girls.

Another strategy for dealing with extraneous variables is to incorporate blocking into the design. In the case of gender, we could create two blocks, one consisting of girls and one consisting of boys. Then, once the blocks are formed, we would randomly assign the girls to the two treatments and randomly assign the boys to the two treatments. In the actual study, the group of 128 children included 81 girls and 47 boys. A diagram that shows the structure of an experiment that includes blocking using gender is shown in Figure 2.9.

---

**FIGURE 2.9**

Diagram for the experiment of Example 2.9 using gender to form blocks.

---

Copyright 2011 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).

Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.
When blocking is used, the design is called a **randomized block design**. Note that one difference between the diagram that describes the experiment in which blocking is used (Figure 2.9) and the diagram of the original experiment (Figure 2.5) is at what point the random assignment occurs. *When blocking is incorporated in an experiment, the random assignment to treatments occurs after the blocks have been formed and is done separately for each block.*

Before proceeding with an experiment, you should be able to give a satisfactory answer to each of the following 10 questions.

1. What is the research question that data from the experiment will be used to answer?
2. What is the response variable?
3. How will the values of the response variable be determined?
4. What are the explanatory variables for the experiment?
5. For each explanatory variable, how many different values are there, and what are these values?
6. What are the treatments for the experiment?
7. What extraneous variables might influence the response?
8. How does the design incorporate random assignment of subjects to treatments (or treatments to subjects) or random assignment of treatments to trials?
9. For each extraneous variable listed in Question 7, how does the design protect against its potential influence on the response through blocking, direct control, or random assignment?
10. Will you be able to answer the research question using the data collected in this experiment?

**EXERCISES 2.33 - 2.47**

2.33 The head of the quality control department at a printing company would like to carry out an experiment to determine which of three different glues results in the greatest binding strength. Although they are not of interest in the current investigation, other factors thought to affect binding strength are the number of pages in the book and whether the book is being bound as a paperback or a hardback.

a. What is the response variable in this experiment?

b. What explanatory variable will determine the experimental conditions?

c. What two extraneous variables are mentioned in the problem description? Are there other extraneous variables that should be considered?

2.34 A study of college students showed a temporary gain of up to 9 IQ points after listening to a Mozart piano sonata. This conclusion, dubbed the Mozart effect, has since been criticized by a number of researchers who have been unable to confirm the result in similar studies. Suppose that you wanted to see whether there is a Mozart effect for students at your school.

a. Describe how you might design an experiment for this purpose.

b. Does your experimental design include direct control of any extraneous variables? Explain.

c. Does your experimental design use blocking? Explain why you did or did not include blocking in your design.

d. What role does random assignment play in your design?
2.35 The following is from an article titled "After the Workout, Got Chocolate Milk?" that appeared in the Chicago Tribune (January 18, 2005):

Researchers at Indiana University at Bloomington have found that chocolate milk effectively helps athletes recover from an intense workout. They had nine cyclists bike, rest four hours, then bike again, three separate times. After each workout, the cyclists downed chocolate milk or energy drinks Gatorade or Endurox (two to three glasses per hour); then, in the second workout of each set, they cycled to exhaustion. When they drank chocolate milk, the amount of time they could cycle until they were exhausted was similar to when they drank Gatorade and longer than when they drank Endurox.

The article is not explicit about this, but in order for this to have been a well-designed experiment, it must have incorporated random assignment. Briefly explain where the researcher would have needed to use random assign in order for the conclusion of the experiment to be valid.

2.36 The report "Comparative Study of Two Computer Mouse Designs" (Cornell Human Factors Laboratory Technical Report RP7992) included the following description of the subjects used in an experiment:

Twenty-four Cornell University students and staff (12 males and 12 females) volunteered to participate in the study. Three groups of 4 men and 4 women were selected by their stature to represent the 5th percentile (female 152.1 ± 0.3 cm, male 164.1 ± 0.4 cm), 50th percentile (female 162.4 ± 0.1 cm, male 174.1 ± 0.7 cm), and 95th percentile (female 171.9 ± 0.2 cm, male 185.7 ± 0.6 cm) ranges . . . All subjects reported using their right hand to operate a computer mouse.

This experimental design incorporated direct control and blocking.

a. Are the potential effects of the extraneous variable stature (height) addressed by blocking or direct control?

b. Whether the right or left hand is used to operate the mouse was considered to be an extraneous variable. Are the potential effects of this variable addressed by blocking or direct control?

2.37 The Institute of Psychiatry at Kings College London found that dealing with “infomania” has a temporary, but significant derogatory effect on IQ (Discover, November 2005). In this experiment, researchers videoed volunteers into two groups. Each subject tool IQ test. One group had to check e-mail and respond instant messages while taking the test, and the sec group took the test without any distraction. The tracted group had an average score that was 10 pc lower than the average for the control group. Ex: why is it important that the researchers created the experimental groups in this study by using ran assignment.

2.38 In an experiment to compare two different surgical procedures for hernia repair ("A Single-Blinded, Randomized Comparison of Laparoscopic Versus Open Hernia Repair in Children," Pediatrics [2009]: 332–336), 89 children were assigned at random to one of two surgical methods. The researchers relied on the dom assignment of subjects to treatments to create comparable groups with respect to extraneous variables they did not control. One such extraneous variable age. After random assignment to treatments, the researchers looked at the age distribution of the children each of the two experimental groups (laparoscopic re (LR) and open repair (OR)). The accompanying figu from the paper.

Based on this figure, has the random assignment of subjects to experimental groups been successful in creating groups that are similar with respect to the ages of children in the groups? Explain.
2.39 In many digital environments, users are allowed to choose how they are represented visually online. Does how people are represented online affect online behavior? This question was examined by the authors of the paper "The Proteus Effect: The Effect of Transformed Self-Representation on Behavior" (Human Communication Research [2007]: 271–290). Participants were randomly assigned either an attractive avatar (a graphical image that represents a person) to represent them or an unattractive avatar.

a. The researchers concluded that when interacting with a person of the opposite gender in an online virtual environment, those assigned an attractive avatar moved significantly closer to the other person than those who had been assigned an unattractive avatar. This difference was attributed to the attractiveness of the avatar. Explain why the researchers would not have been able to reach this conclusion if participants had been allowed to choose one of the two avatars (attractive, unattractive) to represent them online.

b. Construct a diagram to represent the underlying structure of this experiment.

2.40 To examine the effect of exercise on body composition, healthy women age 35 to 50 were classified as either active (9 hours or more of physical activity per week) or sedentary ("Effects of Habitual Physical Activity on the Resting Metabolic Rates and Body Composition of Women aged 35 to 50 Years.” Journal of the American Dietetic Association [2001]: 1181–1191). Percent body fat was measured and the researchers found that percent body fat was significantly lower for women who were active than for sedentary women.

a. Is the study described an experiment? If so, what are the explanatory variable and the response variable? If not, explain why it is not an experiment.

b. From this study alone, is it reasonable to conclude that physical activity is the cause of the observed difference in body fat percentage? Justify your answer.

2.41 Does playing action video games provide more than just entertainment? The authors of the paper "Action-Video-Game Experience Alters the Spatial Resolution of Vision" (Psychological Science [2007]: 88–94) concluded that spatial resolution, an important aspect of vision, is improved by playing action video games. They based this conclusion on data from an experiment in which 32 volunteers who had not played action video games were “equally and randomly divided between the experimental and control groups.” Subjects in each group played a video game for 30 hours over a period of 6 weeks. Those in the experimental group played Unreal Tournament 2004, an action video game. Those in the control group played the game Tetris, a game that does not require the user to process multiple objects at once. Explain why the random assignment to the two groups is an important aspect of this experiment.

2.42 Construct a diagram to represent the subliminal messages experiment of Example 2.5.

2.43 Construct a diagram to represent the gasoline additive experiment described on page 52.

2.44 An advertisement for a sweatshirt that appeared in SkyMall Magazine (a catalog distributed by some airlines) stated the following: “This is not your ordinary hoody! Why? Fact: Research shows that written words on containers of water can influence the water’s structure for better or worse depending on the nature and intent of the word. Fact: The human body is 70% water. What if positive words were printed on the inside of your clothing?” For only $79, you could purchase a hooded sweatshirt that had over 200 positive words (such as hope, gratitude, courage and love) in 15 different languages printed on the inside of the sweatshirt so that you could benefit from being surrounded by these positive words. The reference to the “fact” that written words on containers of water can influence the water’s structure appears to be based on the work of Dr. Masaru Emoto who typed words on paper, pasted the words on bottles of water, and observed how the water reacted to the words by seeing what kind of crystals were formed in the water. He describes several of his experiments in his self-published book, The Message from Water. If you were going to interview Dr. Emoto, what questions would you want to ask him about his experiment?

2.45 An experiment was carried out to assess the effect of Sweet Talk, a text messaging support system for patients with diabetes ("A Randomized Controlled Trial of Sweet Talk,” Diabetic Medicine [2006]: 1332–1338). Participants in the experiment were 92 patients, age 8 to 18, with type I diabetes who had been on conventional insulin treatment for at least one year. Participants were assigned at random to one of three experimental groups:

- Group 1: continued conventional insulin therapy
- Group 2: continued conventional insulin therapy with Sweet Talk support
- Group 3: followed a new intensive insulin therapy with Sweet Talk support

Bold exercises answered in back    ● Data set available online    ✦ Video Solution available
One response variable was a measure of glucose concentration in the blood. There was no significant difference in glucose concentration between groups 1 and 2, but group 3 showed a significant improvement in this measure compared to groups 1 and 2.

a. Explain why it is not reasonable to attribute the observed improvement in group 3 compared to group 1 to the use of Sweet Talk, even though subjects were randomly assigned to the three experimental groups.

b. How would you modify this experiment so that you could tell if improvement in glucose concentration was attributable to the intensive insulin therapy, the use of Sweet Talk, or a combination of the two?

c. Draw a diagram showing the structure of the modified experiment from Part (b).

2.46 The Pew Research Center conducted a study of gender bias. The report “Men or Women: Who is the Better Leader? A Paradox in Public Attitudes” (www.pewsocialtrends.org, August 28, 2008) describes how the study was conducted:

In the experiment, two separate random samples of more than 1000 registered voters were asked to read a profile sent to them online of a hypothetical candidate for U.S. Congress in their district. One random sample of 1161 respondents read a profile of Ann Clark, described as a lawyer, a churchgoer, a member of the local Chamber of Commerce, an environmentalist and a member of the same party as the survey respondent. They were then asked what they liked and didn’t like about her, whether they considered her qualified and whether they were inclined to vote for her. There was no indication that this was a survey about gender or gender bias. A second random sample of 1139 registered voters was asked to read a profile of Andrew Clark who—except for his gender—was identical in every way to Ann Clark. These respondents were then asked the same questions.

a. What are the two treatments in this experiment?

b. What are the response variables in this experiment?

c. Explain why “taking two separate random samples” has the same benefits as random assignment to two treatments in this experiment.

2.47 Red wine contains flavonol, an antioxidant thought to have beneficial health effects. But to have an effect, the antioxidant must be absorbed into the blood. The article “Red Wine is a Poor Source of Bioavailable Flavonols in Men” (The Journal of Nutrition [2001]: 745–748) describes a study to investigate three sources of dietary flavonol—red wine, yellow onions, and black tea—to determine the effect of source on absorption.

The article included the following statement:

We recruited subjects via posters and local newspapers. To ensure that subjects could tolerate the alcohol in the wine, we only allowed men with a consumption of at least seven drinks per week to participate ... Throughout the study, the subjects consumed a diet that was low in flavonols.

a. What are the three treatments in this experiment?

b. What is the response variable?

c. What are three extraneous variables that the researchers chose to control in the experiment?

2.4 More on Experimental Design

The previous section covered basic principles for designing simple comparative experiments—control, blocking, random assignment, and replication. The goal of an experimental design is to provide a method of data collection that (1) minimizes extraneous sources of variability in the response so that any differences in response for various experimental conditions can be more easily assessed and (2) creates experimental groups that are similar with respect to extraneous variables that cannot be controlled either directly or through blocking.

In this section, we look at some additional considerations that you may need to think about when planning an experiment.
Use of a Control Group

If the purpose of an experiment is to determine whether some treatment has an effect, it is important to include an experimental group that does not receive the treatment. Such a group is called a **control group**. The use of a control group allows the experimenter to assess how the response variable behaves when the treatment is not used. This provides a baseline against which the treatment groups can be compared to determine whether the treatment had an effect.

**EXAMPLE 2.10 Comparing Gasoline Additives**

Suppose that an engineer wants to know whether a gasoline additive increases fuel efficiency (miles per gallon). Such an experiment might use a single car (to eliminate car-to-car variability) and a sequence of trials in which 1 gallon of gas is put in an empty tank, the car is driven around a racetrack at a constant speed, and the distance traveled on the gallon of gas is recorded.

To determine whether the additive increases gas mileage, it would be necessary to include a control group of trials in which distance traveled was measured when gasoline without the additive was used. The trials would be assigned **at random** to one of the two experimental conditions (additive or no additive).

Even though this experiment consists of a sequence of trials all with the same car, random assignment of trials to experimental conditions is still important because there will always be uncontrolled variability. For example, temperature or other environmental conditions might change over the sequence of trials, the physical condition of the car might change slightly from one trial to another, and so on. Random assignment of experimental conditions to trials will tend to even out the effects of these uncontrollable factors.

Although we usually think of a control group as one that receives no treatment, in experiments designed to compare a new treatment to an existing standard treatment, the term control group is sometimes also used to describe the group that receives the current standard treatment.

Not all experiments require the use of a control group. For example, many experiments are designed to compare two or more conditions—an experiment to compare density for three different formulations of bar soap or an experiment to determine how oven temperature affects the cooking time of a particular type of cake. However, sometimes a control group is included even when the ultimate goal is to compare two or more different treatments. An experiment with two treatments and no control group might allow us to determine whether there is a difference between the two treatments and even to assess the magnitude of the difference if one exists, but it would not allow us to assess the individual effect of either treatment. For example, without a control group, we might be able to say that there is no difference in the increase in mileage for two different gasoline additives, but we would not be able to tell if this was because both additives increased gas mileage by a similar amount or because neither additive had any effect on gas mileage.

Use of a Placebo

In experiments that use human subjects, use of a control group may not be enough to determine whether a treatment really does have an effect. People sometimes respond merely to the power of suggestion! For example, suppose a study designed to determine
whether a particular herbal supplement is effective in promoting weight loss uses an experimental group that takes the herbal supplement and a control group that takes nothing. It is possible that those who take the herbal supplement and believe that they are taking something that will help them to lose weight may be more motivated and unconsciously change their eating behavior or activity level, resulting in weight loss.

Although there is debate about the degree to which people respond, many studies have shown that people sometimes respond to treatments with no active ingredient and that they often report that such “treatments” relieve pain or reduce symptoms. So, if an experiment is to enable researchers to determine whether a treatment has an effect, comparing a treatment group to a control group may not be enough. To address the problem, many experiments use what is called a placebo.

**DEFINITION**

A **placebo** is something that is identical (in appearance, taste, feel, etc.) to the treatment received by the treatment group, except that it contains no active ingredients.

For example, in the herbal supplement experiment, rather than using a control group that received no treatment, the researchers might want to include a placebo group. Individuals in the placebo group would take a pill that looked just like the herbal supplement but did not contain the herb or any other active ingredient. As long as the subjects did not know whether they were taking the herb or the placebo, the placebo group would provide a better basis for comparison and would allow the researchers to determine whether the herbal supplement had any real effect over and above the “placebo effect.”

**Single-Blind and Double-Blind Experiments**

Because people often have their own personal beliefs about the effectiveness of various treatments, it is desirable to conduct experiments in such a way that subjects do not know what treatment they are receiving. For example, in an experiment comparing four different doses of a medication for relief of headache pain, someone who knows that he is receiving the medication at its highest dose may be subconsciously influenced to report a greater degree of headache pain reduction. By ensuring that subjects are not aware of which treatment they receive, we can prevent the subjects’ personal perceptions from influencing the response.

An experiment in which subjects do not know what treatment they have received is described as **single-blind**. Of course, not all experiments can be made single-blind. For example, in an experiment to compare the effect of two different types of exercise on blood pressure, it is not possible for participants to be unaware of whether they are in the swimming group or the jogging group! However, when it is possible, “blinding” the subjects in an experiment is generally a good strategy.

In some experiments, someone other than the subject is responsible for measuring the response. To ensure that the person measuring the response does not let personal beliefs influence the way in which the response is recorded, the researchers should make sure that the measurer does not know which treatment was given to particular individual. For example, in a medical experiment to determine whether a new vaccine reduces the risk of getting the flu, doctors must decide whether a particular individual who is not feeling well actually has the flu or some other unrelated illness. If the doctor knew that a participant with flu-like symptoms had received the new flu vaccine, she might be less likely to determine that the participant had the flu and more likely to interpret the symptoms as being the result of some other illness.
There are two ways in which blinding might occur in an experiment. One involves blinding the subjects, and the other involves blinding the individuals who measure the response. If subjects do not know which treatment was received and those measuring the response do not know which treatment was given to which subject, the experiment is described as double-blind. If only one of the two types of blinding is present, the experiment is single-blind.

**DEFINITION**

A **double-blind** experiment is one in which neither the subjects nor the individuals who measure the response know which treatment was received.

A **single-blind** experiment is one in which the subjects do not know which treatment was received but the individuals measuring the response do know which treatment was received, or one in which the subjects do know which treatment was received but the individuals measuring the response do not know which treatment was received.

**Experimental Units and Replication**

An **experimental unit** is the smallest unit to which a treatment is applied. In the language of experimental design, treatments are assigned at random to experimental units, and replication means that each treatment is applied to more than one experimental unit.

Replication is necessary for random assignment to be an effective way to create similar experimental groups and to get a sense of the variability in the values of the response for individuals who receive the same treatment. As we will see in Chapters 9–15, this enables us to use statistical methods to decide whether differences in the responses in different treatment groups can be attributed to the treatment received or whether they can be explained by chance variation (the natural variability seen in the responses to a single treatment).

Be careful when designing an experiment to ensure that there is replication. For example, suppose that children in two third-grade classes are available to participate in an experiment to compare two different methods for teaching arithmetic. It might at first seem reasonable to select one class at random to use one method and then assign the other method to the remaining class. But what are the experimental units here? If treatments are randomly assigned to classes, classes are the experimental units. Because only one class is assigned to each treatment, this is an experiment with no replication, even though there are many children in each class. We would not be able to determine whether there was a difference between the two methods based on data from this experiment, because we would have only one observation per treatment.

One last note on replication: Do not confuse replication in an experimental design with replicating an experiment. Replicating an experiment means conducting a new experiment using the same experimental design as a previous experiment; it is a way of confirming conclusions based on a previous experiment, but it does not eliminate the need for replication in each of the individual experiments themselves.

**Using Volunteers as Subjects in an Experiment**

Although the use of volunteers in a study that involves collecting data through sampling is never a good idea, it is a common practice to use volunteers as subjects in an experiment. Even though the use of volunteers limits the researcher’s ability to generalize to a larger population, random assignment of the volunteers to treatments should result in comparable groups, and so treatment effects can still be assessed.
2.48 Explain why some studies include both a control group and a placebo treatment. What additional comparisons are possible if both a control group and a placebo group are included?

2.49 Explain why blinding is a reasonable strategy in many experiments.

2.50 Give an example of an experiment for each of the following:
   a. Single-blind experiment with the subjects blinded
   b. Single-blind experiment with the individuals measuring the response blinded
   c. Double-blind experiment
   d. An experiment for which it is not possible to blind the subjects

2.51 Swedish researchers concluded that viewing and discussing art soothes the soul and helps relieve medical conditions such as high blood pressure and constipation (AFP International News Agency, October 14, 2005). This conclusion was based on a study in which 20 elderly women gathered once a week to discuss different works of art. The study also included a control group of 20 elderly women who met once a week to discuss their hobbies and interests. At the end of 4 months, the art discussion group was found to have a more positive attitude, to have lower blood pressure, and to use fewer laxatives than the control group.
   a. Why would it be important to determine if the researchers assigned the women participating in the study at random to one of the two groups?
   b. Explain why you think that the researchers included a control group in this study.

2.52 In an experiment to compare two different surgical procedures for hernia repair (“A Single-Blinded, Randomized Comparison of Laparoscopic Versus Open Hernia Repair in Children,” Pediatrics [2009]: 332–336), 89 children were assigned at random to one of the two surgical methods. The methods studied were laparoscopic repair and open repair. In laparoscopic repair, three small incisions are made and the surgeon works through these incisions with the aid of a small camera that is inserted through one of the incisions. In the open repair, a larger incision is used to open the abdomen. One of the response variables in this study was the amount of medication that was given after the surgery for the control of pain and nausea. The paper states “For postoperative pain, rescue fentanyl (1 µg/kg) and ondansetron (0.1 mg/kg) were given as judged necessary by the attending nurse blinded to the operative approach.”
   a. Why do you think it was important that the nurse who administered the medications did not know which type of surgery was performed?
   b. Explain why it was not possible for this experiment to be double-blind.

2.53 The article “Placebos Are Getting More Effective. Drug Makers Are Desperate to Know Why” (Wired Magazine, August 8, 2009) states that “according to research, the color of a tablet can boost the effectiveness even of genuine meds—or help convince a patient that a placebo is a potent remedy.” Describe how you would design an experiment to investigate if adding a color to Tylenol tablets would result in greater perceived pain relief. Be sure to address how you would select subjects, how you would measure pain relief, what color you would use, and whether or not you would include a control group in your experiment.

2.54 A novel alternative medical treatment for heart attacks seeds the damaged heart muscle with cells from the patient’s thigh muscle (“Doctors Mend Damaged Hearts with Cells from Muscles,” San Luis Obispo Tribune, November 18, 2002). Doctor Dib from the area Heart Institute evaluated the approach on 16 patients with severe heart failure. The article states “ordinarily, the heart pushes out more than half its blood with each beat. Dib’s patients had such severe heart failure that their hearts pumped just 23 percent of blood per beat. After bypass surgery and cell injections, this improved to 36 percent, although it was impossible to say if much, if any, of the new strength resulted from extra cells.”
   a. Explain why it is not reasonable to generalize to the population of all heart attack victims based on data from these 16 patients.
   b. Explain why it is not possible to say whether an apparent improvement was due to the cell injections, based on the results of this study.
   c. Describe a design for an experiment that would allow researchers to determine whether bypass surgery plus cell injections was more effective than by surgery alone.
2.55 The article “Doctor Dogs Diagnose Cancer by Sniffing It Out” (Knight Ridder Newspapers, January 9, 2006) reports the results of an experiment described in the journal Integrative Cancer Therapies. In this experiment, dogs were trained to distinguish between people with breast and lung cancer and people without cancer by sniffing exhaled breath. Dogs were trained to lay down if they detected cancer in a breath sample. After training, dogs’ ability to detect cancer was tested using breath samples from people whose breath had not been used in training the dogs. The paper states “The researchers blinded both the dog handlers and the experimental observers to the identity of the breath samples.” Explain why this blinding is an important aspect of the design of this experiment.

2.56 An experiment to evaluate whether vitamins can help prevent recurrence of blocked arteries in patients who have had surgery to clear blocked arteries was described in the article “Vitamins Found to Help Prevent Blocked Arteries” (Associated Press, September 1, 2002). The study involved 205 patients who were given either a treatment consisting of a combination of folic acid, vitamin B12, and vitamin B6 or a placebo for 6 months.

a. Explain why a placebo group was used in this experiment.

b. Explain why it would be important for the researchers to have assigned the 205 subjects to the two groups (vitamin and placebo) at random.

c. Do you think it is appropriate to generalize the results of this experiment to the population of all patients who have undergone surgery to clear blocked arteries? Explain.

2.57 Pismo Beach, California, has an annual clam festival that includes a clam chowder contest. Judges rate clam chowders from local restaurants, and the judging is done in such a way that the judges are not aware of which chowder is from which restaurant. One year, much to the dismay of the seafood restaurants on the waterfront, Denny’s chowder was declared the winner! (When asked what the ingredients were, the cook at Denny’s said he wasn’t sure—he just had to add the right amount of nondairy creamer to the soup stock that he got from Denny’s distribution center!)

a. Do you think that Denny’s chowder would have won the contest if the judging had not been “blind?” Explain.

b. Although this was not an experiment, your answer to Part (a) helps to explain why those measuring the response in an experiment are often blinded. Using your answer in Part (a), explain why experiments are often blinded in this way.

2.58 The San Luis Obispo Tribune (May 7, 2002) reported that “a new analysis has found that in the majority of trials conducted by drug companies in recent decades, sugar pills have done as well as—or better than—antidepressants.” What effect is being described here? What does this imply about the design of experiments with a goal of evaluating the effectiveness of a new medication?

2.59 The article “A Debate in the Dentist’s Chair” (San Luis Obispo Tribune, January 28, 2000) described an ongoing debate over whether newer resin fillings are a better alternative to the more traditional silver amalgam fillings. Because amalgam fillings contain mercury, there is concern that they could be mildly toxic and prove to be a health risk to those with some types of immune and kidney disorders. One experiment described in the article used sheep as subjects and reported that sheep treated with amalgam fillings had impaired kidney function.

a. In the experiment, a control group of sheep that received no fillings was used but there was no placebo group. Explain why it is not necessary to have a placebo group in this experiment.

b. The experiment compared only an amalgam filling treatment group to a control group. What would be the benefit of also including a resin filling treatment group in the experiment?

c. Why do you think the experimenters used sheep rather than human subjects?

2.5 More on Observational Studies: Designing Surveys (Optional)

Designing an observational study to compare two populations on the basis of some easily measured characteristic is relatively straightforward, with attention focusing on choosing a reasonable method of sample selection. However, many observational
studies attempt to measure personal opinion or attitudes using responses to a survey. In such studies, both the sampling method and the design of the survey itself are critical to obtaining reliable information.

At first glance it might seem that a survey is a simple method for acquiring information. However, it turns out that designing and administering a survey is no easy task. Great care must be taken in order to obtain good information from a survey.

Survey Basics

A survey is a voluntary encounter between strangers in which an interviewer seeks information from a respondent by engaging in a special type of conversation. The conversation might take place in person, over the telephone, or even in the form of a written questionnaire, and it is quite different from usual social conversations. The interviewer and the respondent have certain roles and responsibilities. The interviewer gets to decide what is relevant to the conversation and may ask questions—possibly personal or even embarrassing questions. The respondent, in turn, may refuse to participate in the conversation and may refuse to answer any particular question. But having agreed to participate in the survey, the respondent is responsible for answering the questions truthfully. Let’s consider the situation of the respondent.

The Respondent’s Tasks

Understanding of the survey process has been improved in the past two decades by contributions from the field of psychology, but there is still much uncertainty about how people respond to survey questions. Survey researchers and psychologists generally agree that the respondent is confronted with a sequence of tasks when asked a question: comprehension of the question, retrieval of information from memory, and reporting the response.

Task 1: Comprehension

Comprehension is the single most important task facing the respondent, and fortunately it is the characteristic of a survey question that is most easily controlled by the question writer. Understandable directions and questions are characterized by (1) a vocabulary appropriate to the population of interest, (2) simple sentence structure, and (3) little or no ambiguity. Vocabulary is often a problem. As a rule, it is best to use the simplest possible word that can be used without sacrificing clear meaning.

Simple sentence structure also makes it easier for the respondent to understand the question. A famous example of difficult syntax occurred in 1993 when the Roper organization created a survey related to the Holocaust. One question in this survey was:

“Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?”

The question has a complicated structure and a double negative—“impossible never happened”—that could lead respondents to give an answer opposite to what they actually believed. The question was rewritten and given a year later in an otherwise unchanged survey:

“Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?”

This question wording is much clearer, and in fact the respondents’ answers were quite different, as shown in the following table (the “unsure” and “no opinion” percentages have been omitted):
It is also important to filter out ambiguity in questions. Even the most innocent and seemingly clear questions can have a number of possible interpretations. For example, suppose that you are asked, “When did you move to Cedar Rapids?” This would seem to be an unambiguous question, but some possible answers might be (1) “In 1971,” (2) “When I was 23,” and (3) “In the summer.” The respondent must decide which of these three answers, if any, is the appropriate response. It may be possible to lessen the ambiguity with more precise questions:

1. In what year did you move to Cedar Rapids?
2. How old were you when you moved to Cedar Rapids?
3. In what season of the year did you move to Cedar Rapids?

One way to find out whether or not a question is ambiguous is to field-test the question and to ask the respondents if they were unsure how to answer a question.

Ambiguity can also arise from the placement of questions as well as from their phrasing. Here is an example of ambiguity uncovered when the order of two questions differed in two versions of a survey on happiness. The questions were

1. Taken altogether, how would you say things are these days: Would you say that you are very happy, pretty happy, or not too happy?

2. Taking things altogether, how would you describe your marriage: Would you say that your marriage is very happy, pretty happy, or not too happy?

The proportions of responses to the general happiness question differed for the different question orders, as follows:

<table>
<thead>
<tr>
<th>General Asked First</th>
<th>General Asked Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very happy</td>
<td>52.4%</td>
</tr>
<tr>
<td>Pretty happy</td>
<td>44.2%</td>
</tr>
<tr>
<td>Not too happy</td>
<td>3.4%</td>
</tr>
<tr>
<td></td>
<td>38.1%</td>
</tr>
<tr>
<td></td>
<td>52.8%</td>
</tr>
<tr>
<td></td>
<td>9.1%</td>
</tr>
</tbody>
</table>

If the goal in this survey was to estimate the proportion of the population that is generally happy, these numbers are quite troubling—they cannot both be right! What seems to have happened is that Question 1 was interpreted differently depending on whether it was asked first or second. When the general happiness question was asked after the marital happiness question, the respondents apparently interpreted it to be asking about their happiness in all aspects of their lives except their marriage. This was a reasonable interpretation, given that they had just been asked about their marital happiness, but it is a different interpretation than when the general happiness question was asked first. The troubling lesson here is that even carefully worded questions can have different interpretations in the context of the rest of the survey.
Task 2: Retrieval from Memory  Retrieving relevant information from memory can be a challenging task, and it is not always an easy task, and it is not a problem limited to questions of fact. For example, consider this seemingly elementary “factual” question:

How many times in the past 5 years did you visit your dentist’s office?

   a. 0 times
   b. 1–5 times
   c. 6–10 times
   d. 11–15 times
   e. more than 15 times

It is unlikely that many people will remember with clarity every single visit to the dentist in the past 5 years. But generally, people will respond to such a question with answers consistent with the memories and facts they are able to reconstruct given time they have to respond to the question. An individual may, for example, have a sense that he usually makes about two trips a year to the dentist’s office, so he extrapolates the typical year and gets 10 times in 5 years. Then there may be three particularly memorable visits, say, for a root canal in the middle of winter. Thus, the recollection is now 13, and the respondent will choose Answer (d), 11–15 times. Perhaps not exactly correct, but the best that can be reported under the circumstances.

What are the implications of this relatively fuzzy memory for those who construct surveys about facts? First, the investigator should understand that most factual answers are going to be approximations of the truth. Second, events closer to the time of a survey are easier to recall.

Attitude and opinion questions can also be affected in significant ways by the respondent’s memory of recently asked questions. For example, one study asked survey question asking respondents their opinion about how much they followed politics. When that question was preceded by a factual question asking whether they knew the name of the congressional representative from their district, the percentage reported they follow politics “now and then” or “hardly ever” jumped from 21% to 39%! Respondents apparently concluded that, because they didn’t know the answer to the previous knowledge question, they must not follow politics as much as they might have thought otherwise. In a survey that asks for an opinion about the degree to which the respondent believes drilling for oil should be permitted in national parks, the response might be different if the question is preceded by questions about the high price of gasoline than if the question is preceded by questions about the environment.

Task 3: Reporting the Response  The task of formulating and reporting a response can be influenced by the social aspects of the survey conversation. In general, if a respondent agrees to take a survey, he or she will be motivated to answer truthfully. Therefore, if questions are not too difficult (taxing the respondent’s knowledge or memory) and if there are not too many questions (taxing the respondent’s patience), the answers to questions will be reasonably accurate. However, it is also true that the respondents often wish to present themselves in a favorable light. This desire leads to what is known as a social desirability bias. Sometimes this bias is a response to the particular wording in a question. In the following questions were analyzed in two different forms of a survey (emphasis added):

1. Do you think the United States should forbid public speeches against democracy?
2. Do you think the United States should allow public speeches against democracy?

It would seem logical that these questions are opposites and that the proportion who would not allow public speeches against democracy should be equal to the proportion who would forbid public speeches against democracy. But only 45% of those respondents offering an opinion on Question 1 thought the United States should “forbid”
whereas 75% of the respondents offering an opinion on Question 2 thought the United States should “not allow” public speeches against democracy. Most likely, respondents reacted negatively to the word *forbid*, as forbidding something sounds much harsher than not allowing it.

Some survey questions may be sensitive or threatening, such as questions about sex, drugs, or potentially illegal behavior. In this situation, a respondent not only will want to present a positive image but also will certainly think twice about admitting illegal behavior! In such cases, the respondent may shade the actual truth or may even lie about particular activities and behaviors. In addition, the tendency toward positive presentation is not limited to obviously sensitive questions. For example, consider the question about general happiness previously described. Several investigators have reported higher happiness scores in face-to-face interviews than in responses to a mailed questionnaire. Presumably, a happy face presents a more positive image of the respondent to the interviewer. On the other hand, if the interviewer was a clearly unhappy person, a respondent might shade answers to the less happy side of the scale, perhaps thinking that it is inappropriate to report happiness in such a situation.

It is clear that constructing surveys and writing survey questions can be a daunting task. Keep in mind the following three things:

1. Questions should be understandable by the individuals in the population being surveyed. Vocabulary should be at an appropriate level, and sentence structure should be simple.
2. Questions should, as much as possible, recognize that human memory is fickle. Questions that are specific will aid the respondent by providing better memory cues. The limitations of memory should be kept in mind when interpreting the respondent’s answers.
3. As much as possible, questions should not create opportunities for the respondent to feel threatened or embarrassed. In such cases respondents may introduce a social desirability bias, the degree of which is unknown to the interviewer. This can compromise conclusions drawn from the survey data.

Constructing good surveys is a difficult task, and we have given only a brief introduction to this topic. For a more comprehensive treatment, we recommend the book by Sudman and Bradburn listed in the references in the back of the book.

---

**EXERCISES 2.60 - 2.65**

2.60 A tropical forest survey conducted by Conservation International included the following statements in the material that accompanied the survey:

“A massive change is burning its way through the earth’s environment.”

“The band of tropical forests that encircle the earth is being cut and burned to the ground at an alarming rate.”

“Never in history has mankind inflicted such sweeping changes on our planet as the clearing of rain forest taking place right now!”

The survey that followed included the questions given in Parts (a)–(d) below. For each of these questions, identify a word or phrase that might affect the response and possibly bias the results of any analysis of the responses.

a. “Did you know that the world’s tropical forests are being destroyed at the rate of 80 acres per minute?”

b. “Considering what you know about vanishing tropical forests, how would you rate the problem?”

c. “Do you think we have an obligation to prevent the man-made extinction of animal and plant species?”

d. “Based on what you know now, do you think there is a link between the destruction of tropical forests and changes in the earth’s atmosphere?”

**Bold** exercises answered in back

● Data set available online

★ Video Solution available

---

Copyright 2011 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.
2.61 Fast-paced lifestyles, in which students balance the requirements of school, after-school activities, and jobs, are thought by some to lead to reduced sleep. Suppose that you are assigned the task of designing a survey that will provide answers to the accompanying questions. Write a set of survey questions that might be used. In some cases, you may need to write more than one question to adequately address a particular issue. For example, responses might be different for weekends and school nights. You may also have to define some terms to make the questions understandable to the target audience, which is adolescents.

Topics to be addressed:
How much sleep do the respondents get? Is this enough sleep?
Does sleepiness interfere with schoolwork?
If they could change the starting and ending times of the school day, what would they suggest? (Sorry, they cannot reduce the total time spent in school during the day!)

2.62 Asthma is a chronic lung condition characterized by difficulty in breathing. Some studies have suggested that asthma may be related to childhood exposure to some animals, especially dogs and cats, during the first year of life ("Exposure to Dogs and Cats in the First Year of Life and Risk of Allergic Sensitization at 6 to 7 Years of Age." Journal of the American Medical Association [2002]: 963–972). Some environmental factors that trigger an asthmatic response are (1) cold air, (2) dust, (3) strong fumes, and (4) inhaled irritants.

a. Write a set of questions that could be used in a survey to be given to parents of young children suffering from asthma. The survey should include questions about the presence of pets in the first year of the child’s life as well as questions about the presence of pets today. Also, the survey should include questions that address the four mentioned household environmental factors.

b. It is generally thought that low-income persons, who tend to be less well educated, have homes in environments where the four environmental factors are present. Mindful of the importance of comprehension, can you improve the questions in Part (a) by making your vocabulary simpler or by changing the wording of the questions?

c. One problem with the pet-related questions is reliance on memory. That is, parents may not act remember when they got their pets. How might check the parents’ memories about these pets?

2.63 In national surveys, parents consistently pair school safety as an important concern. One source of violence in junior high schools is fighting ("Self-Report Characterization of Seventh-Grade Student Fights," Journal of Adolescent Health [1998]: 103–109). To construct a knowledge base about student fights, a school administrator wants to give two surveys to students: one of the two surveys is to be given to the participants, and the other is to be given to students who witnessed the fight. The type of information desired includes (1) the cause of the fight, (2) whether or not fight was a continuation of a previous fight, (3) whether drugs or alcohol was a factor, (4) whether or not the fight was gang related, and (5) the role of bystanders.

a. Write a set of questions that could be used in two surveys. Each question should include a set of possible responses. For each question, indicate whether it would be used on both surveys or just one of the two.

b. How might the tendency toward positive presentation affect the responses of the fighter to survey questions you wrote for Part (a)?

c. How might the tendency toward positive presentation affect the responses of a bystander to the survey questions you wrote for Part (a)?

2.64 Doctors have expressed concern about young women drinking large amounts of soda and about their decreased consumption of milk ("Teenaged Girls, Carbonated Beverage Consumption, and Bone Fractures," Archives of Pediatric and Adolescent Medicine [2001]: 610–613). In parts (a)–(d), construct two questions that might be included in a survey of teenage girls. Each question should include possible responses from which the respondent can select. (Note: The question written are vague. Your task is to clarify the question: use in a survey, not just to change the syntax!)

a. How much “cola” beverage does the respondent consume?

b. How much milk (and milk products) is consumed by the respondent?

c. How physically active is the respondent?

d. What is the respondent’s history of bone fractures?
2.65 A survey described in the paper “The Adolescent Health Review: A Brief Multidimensional Screening Instrument” (Journal of Adolescent Health [2001]:131–139) attempted to address psychosocial factors thought to be of importance in preventive health care for adolescents. For each risk area in the following list, construct a question that would be comprehensible to students in grades 9–12 and that would provide information about the risk factor.

Make your questions multiple-choice, and provide possible responses.

a. Lack of exercise
b. Poor nutrition
c. Emotional distress
d. Sexual activity
e. Cigarette smoking
f. Alcohol use

2.6 Interpreting and Communicating the Results of Statistical Analyses

Statistical studies are conducted to allow investigators to answer questions about characteristics of some population of interest or about the effect of some treatment. Such questions are answered on the basis of data, and how the data are obtained determines the quality of information available and the type of conclusions that can be drawn. As a consequence, when describing a study you have conducted (or when evaluating a published study), you must consider how the data were collected.

The description of the data collection process should make it clear whether the study is an observational study or an experiment. For observational studies, some of the issues that should be addressed are:

1. What is the population of interest? What is the sampled population? Are these two populations the same? If the sampled population is only a subset of the population of interest, **undercoverage** limits our ability to generalize to the population of interest. For example, if the population of interest is all students at a particular university, but the sample is selected from only those students who choose to list their phone number in the campus directory, undercoverage may be a problem. We would need to think carefully about whether it is reasonable to consider the sample as representative of the population of all students at the university. **Overcoverage** results when the sampled population is actually larger than the population of interest. This would be the case if we were interested in the population of all high schools that offer Advanced Placement (AP) Statistics but sampled from a list of all schools that offered an AP class in any subject. Both undercoverage and overcoverage can be problematic.

2. How were the individuals or objects in the sample actually selected? A description of the sampling method helps the reader to make judgments about whether the sample can reasonably be viewed as representative of the population of interest.

3. What are potential sources of bias, and is it likely that any of these will have a substantial effect on the observed results? When describing an observational study, you should acknowledge that you are aware of potential sources of bias and explain any steps that were taken to minimize their effect. For example, in a mail survey, nonresponse can be a problem, but the sampling plan may seek to minimize its effect by offering incentives for participation and by following up one or more times with those who do not respond to the first request. A common misperception is that increasing the sample size is a way to reduce bias in observational studies, but this is not the case. For example, if measurement bias is
present, as in the case of a scale that is not correctly calibrated and tends to weigh too high, taking 1000 measurements rather than 100 measurements cannot correct for the fact that the measured weights will be too large. Similarly, a larger sample size cannot compensate for response bias introduced by a poorly worded question.

For experiments, some of the issues that should be addressed are:

1. What is the role of random assignment? All good experiments use random assignment as a means of coping with the effects of potentially confounding variables that cannot easily be directly controlled. When describing an experimental design, you should be clear about how random assignment (subjects to treatments, or treatments to trials) was incorporated into the design.

2. Were any extraneous variables directly controlled by holding them at fixed values throughout the experiment? If so, which ones and at which values?

3. Was blocking used? If so, how were the blocks created? If an experiment blocking to create groups of homogeneous experimental units, you should describe the criteria used to create the blocks and their rationale. For example, might say something like “Subjects were divided into two blocks—those who exercise regularly and those who do not exercise regularly—because it was believed that exercise status might affect the responses to the diets.”

Because each treatment appears at least once in each block, the block size r be at least as large as the number of treatments. Ideally, the block sizes should be equal to the number of treatments, because this presumably would allow the experimenter to create small groups of extremely homogeneous experimental units. For example, an experiment to compare two methods for teaching calculus to first-year college students, we may want to block on previous mathematics knowledge by using n SAT scores. If 100 students are available as subjects for this experiment, rather than creating two large groups (above-average math SAT score and below-average SAT score), we might want to create 50 blocks of two students each, the first containing the two students with the highest math SAT scores, the second containing two students with the next highest scores, and so on. We would then select one student in each block at random and assign that student to teaching method 1, the other student in the block would be assigned to teaching method 2.

A Word to the Wise: Cautions and Limitations

It is a big mistake to begin collecting data before thinking carefully about research objectives and developing a plan. A poorly designed plan for data collection may result in data that do not enable the researcher to answer key questions of interest or to generalize conclusions based on the data to the desired populations of interest.

Clearly defining the objectives at the outset enables the investigator to determine whether an experiment or an observational study is the best way to proceed. Watch out for the following inappropriate actions:

1. Drawing a cause-and-effect conclusion from an observational study. Don’t this, and don’t believe it when others do it!

2. Generalizing results of an experiment that uses volunteers as subjects to a larger population. This is not sensible without a convincing argument that the group of volunteers can reasonably be considered a representative sample from the population.

3. Generalizing conclusions based on data from a sample to some population of interest. This is sometimes a sensible thing to do, but on other occasions it is
reasonable. Generalizing from a sample to a population is justified only when there is reason to believe that the sample is likely to be representative of the population. This would be the case if the sample was a random sample from the population and there were no major potential sources of bias. If the sample was not selected at random or if potential sources of bias were present, these issues would have to be addressed before a judgment could be made regarding the appropriateness of generalizing the study results.

For example, the Associated Press (January 25, 2003) reported on the high cost of housing in California. The median home price was given for each of the 10 counties in California with the highest home prices. Although these 10 counties are a sample of the counties in California, they were not randomly selected and (because they are the 10 counties with the highest home prices) it would not be reasonable to generalize to all California counties based on data from this sample.

4. Generalizing conclusions based on an observational study that used voluntary response or convenience sampling to a larger population. This is almost never reasonable.

EXERCISES 2.66 - 2.69

2.66 The following paragraph appeared in USA Today (August 6, 2009):

Cement doesn’t hold up to scrutiny

A common treatment that uses medical cement to fix cracks in the spinal bones of elderly people worked no better than a sham treatment, the first rigorous studies of a popular procedure reveal. Pain and disability were virtually the same up to six months later, whether patients had a real treatment or a fake one, shows the research in today’s New England Journal of Medicine. Tens of thousands of Americans each year are treated with bone cement, especially older women with osteoporosis. The researchers said it is yet another example of a procedure coming into wide use before proven safe and effective. Medicare pays $1,500 to $2,100 for the outpatient procedure.

The paper referenced in this paragraph is "A Randomized Trial of Vertebroplasty for Painful Osteoporotic Vertebral Fractures" (New England Journal of Medicine [2009]: 557-568). Obtain a copy of this paper through your university library or your instructor. Read the following sections of the paper: the abstract on page 557; the study design section on page 558; the participants section on pages 558–559; the outcome assessment section on pages 559–560; and the discussion section that begins on page 564.

The summary of this study that appeared in USA Today consisted of just one paragraph. If the newspaper had allowed four paragraphs, other important aspects of the study could have been included. Write a four-paragraph summary that the paper could have used. Remember—you are writing for the USA Today audience, not for the readers of the New England Journal of Medicine!

2.67 The article "Effects of Too Much TV Can Be Undone" (USA Today, October 1, 2007) included the following paragraph:

Researchers at Johns Hopkins Bloomberg School of Public Health report that it’s not only how many hours children spend in front of the TV, but at what age they watch that matters. They analyzed data from a national survey in which parents of 2707 children were interviewed first when the children were 30–33 months old and again when they were 5 1/2, about their TV viewing and their behavior.

a. Is the study described an observational study or an experiment?

b. The article says that data from a sample of 2707 parents were used in the study. What other information about the sample would you want in order to evaluate the study?
c. The actual paper referred to by the USA Today article was “Children’s Television Exposure and Behavioral and Social Outcomes at 5.5 years: Does Timing of Exposure Matter?” (Pediatrics [2007]: 762–769). The paper describes the sample as follows:

The study sample included 2707 children whose mothers completed telephone interviews at both 30 to 33 months and 5.5 years and reported television exposure at both time points. Of those completing both interviews, 41 children (1%) were excluded because of missing data on television exposure at one or both time points. Compared with those enrolled in the HS clinical trial, parents in the study sample were disproportionately older, white, more educated, and married.

The “HS clinical trial” referred to in the excerpt from the paper was a nationally representative sample used in the Healthy Steps for Young Children national evaluation. Based on the above description of the study sample, do you think that it is reasonable to regard the sample as representative of parents of all children at age 5.5 years? Explain.

d. The USA Today article also includes the following summary paragraph:

The study did not examine what the children watched and can’t show TV was the cause of later problems, but it does “tell parents that even if kids are watching TV early in life, and they stop, it could reduce the risk for behavioral and social problems later,” Mistry says.

ACTIVITY 2.1 Facebook Friending

Background: The article “Professors Prefer Face Time to Facebook” appeared in the student newspaper at Cal Poly, San Luis Obispo (Mustang Daily, August 27, 2009). The article examines how professors and students felt about using Facebook as a means of faculty-student communication. The student who wrote this article got mixed opinions when she interviewed students to ask whether they wanted to become Facebook friends with their professors. Two student comments included in the article were

“I think the younger the professor is, the more you can relate to them and the less awkward it would be if you were to become friends on Facebook. The older the professor, you just would have to won ‘Why are they friending me?’” and

“I think becoming friends with professors on Facebook is really awkward. I don’t want them being able to see into my personal life, and frankly, I’m not really interested in what my professors do in their free time.”

Even if the students interviewed had expressed a content opinion, it would still be unreasonable to think represented general student opinion on this issue because only four students were interviewed and it is not clear from the article how these students were selected.

What potentially confounding variable is identified in this passage?

e. The passage in Part (d) says that the study cannot show that TV was the cause of later problems. Is the quote from Kamila Mistry (one of the study authors) in the passage consistent with the statement about the cause? Explain.

2.68 The short article “Developing Science-Based Food and Nutrition Information” (Journal of the American Dietetic Association [2001]: 1144–1145) includes guidelines for evaluating a research paper. Obtain a copy of this paper through your university library or your instructor. Read this article and make a list of questions that be used to evaluate a research study.

2.69 An article titled “I Said, Not While You Study: Why Are You Studying While You Study?” appeared in the Washington Post (September 5, 2006). This provides an example of a reporter summarizing the result of a scientific study in a way that is designed to make it accessible to the newspaper’s readers. You can find the newspaper article online by searching on the title or by going to http://www.washingtonpost.com/wp-dyn/content/article/2006/09/03/AR2006090300.html. The study referenced in the newspaper article published in the Proceedings of the National Academy of Science and can be found at http://www.pnas.content/103/31/11778.full.

Read the newspaper article and then take a look the published paper. Comment on whether you think that the author was successful in communicating findings of the study to the intended audience.
In this activity, you will work with a partner to develop a plan to assess student opinion about being Facebook friends with professors at your school.

1. Suppose you will select a sample of 50 students at your school to participate in a survey. Write one or more questions that you would ask each student in the sample.

2. Discuss with your partner whether you think it would be easy or difficult to obtain a simple random sample of 50 students at your school and to obtain the desired information from all the students selected for the sample. Write a summary of your discussion.

3. With your partner, decide how you might go about selecting a sample of 50 students from your school that reasonably could be considered representative of the population of interest even if it may not be a simple random sample. Write a brief description of your sampling plan, and point out the aspects of your plan that you think make it reasonable to argue that it will be representative.

4. Explain your plan to another pair of students. Ask them to critique your plan. Write a brief summary of the comments you received. Now reverse roles, and provide a critique of the plan devised by the other pair.

5. Based on the feedback you received in Step 4, would you modify your original sampling plan? If not, explain why this is not necessary. If so, describe how the plan would be modified.

**ACTIVITY 2.2 An Experiment to Test for the Stroop Effect**

**Background:** In 1935, John Stroop published the results of his research into how people respond when presented with conflicting signals. Stroop noted that most people are able to read words quickly and that they cannot easily ignore them and focus on other attributes of a printed word, such as text color. For example, consider the following list of words:

```
green  blue  red  blue  yellow  red
```

It is easy to quickly read this list of words. It is also easy to read the words even if the words are printed in color, and even if the text color is different from the color of the word. For example, people can read the words in the list

```
green  blue  red  blue  yellow  red
```
as quickly as they can read the list that isn’t printed in color.

However, Stroop found that if people are asked to name the text colors of the words in the list (red, yellow, blue, green, red, green), it takes them longer. Psychologists believe that this is because the reader has to inhibit a natural response (reading the word) and produce a different response (naming the color of the text).

If Stroop is correct, people should be able to name colors more quickly if they do not have to inhibit the word response, as would be the case if they were shown the following:

```
[ ] [ ] [ ] [ ] [ ] [ ]
```

1. Design an experiment to compare times to identify colors when they appear as text to times to identify colors when there is no need to inhibit a word response. Indicate how random assignment is incorporated into your design. What is your response variable? How will you measure it? How many subjects will you use in your experiment, and how will they be chosen?

2. When you are satisfied with your experimental design, carry out the experiment. You will need to construct your list of colored words and a corresponding list of colored bars to use in the experiment. You will also need to think about how you will implement the random assignment scheme.

3. Summarize the resulting data in a brief report that explains whether your findings are consistent with the Stroop effect.
**ACTIVITY 2.3  McDonald’s and the Next 100 Billion Burgers**

**Background:** The article “Potential Effects of the Next 100 Billion Hamburgers Sold by McDonald’s” (American Journal of Preventative Medicine [2005]: 379–381) estimated that 992.25 million pounds of saturated fat would be consumed as McDonald’s sells its next 100 billion hamburgers. This estimate was based on the assumption that the average weight of a burger sold would be 2.4 oz. This is the average of the weight of a regular hamburger (1.6 oz.) and a Big Mac (3.2 oz.). The authors took this approach because McDonald’s does not publish sales and profits of individual items. Thus, it is not possible to estimate how many of McDonald’s first 100 billion beef burgers sold were 1.6 oz hamburgers, 3.2 oz. Big Macs (introduced in 1968), 4.0 oz. Quarter Pounders (introduced in 1973), or other sandwiches.

This activity can be completed as an individual or team. Your instructor will specify which approach (individual or team) you should use.

1. The authors of the article believe that the use of 2.4 oz. as the average size of a burger sold at McDonald’s is “conservative,” which would result in the estimate of 992.25 million pounds of saturated fat being lower than the actual amount that will be consumed. Explain why the authors’ belief might be justified.

2. Do you think it would be possible to collect data that could lead to an estimate of the average burger size that would be better than 2.4 oz.? If so, explain how you would recommend collecting such data. If not, explain why you think it is not possible.

---

**ACTIVITY 2.4  Video Games and Pain Management**

**Background:** Video games have been used for pain management by doctors and therapists who believe that the attention required to play a video game can distract the player and thereby decrease the sensation of pain. The paper “Video Games and Health” (British Medical Journal [2005]:122–123) states

However, there has been no long term follow-up and no robust randomized controlled trials of such interventions. Whether patients eventually tire of such games is also unclear. Furthermore, it is not known whether any distracting effect depends simply on concentrating on an interactive task or whether the content of games is also an important factor as there have been no controlled trials comparing video games with other distracters. Further research should examine factors within games such as novelty, users’ preferences, and relative levels of challenge and should compare video games with other potentially distracting activities.

1. Working with a partner, select one of the areas of potential research suggested in the passage from paper and formulate a specific question that could be addressed by performing an experiment.

2. Propose an experiment that would provide data to address the question from Step 1. Be specific about how subjects might be selected, what the experimental conditions (treatments) would be, and what response would be measured.

3. At the end of Section 2.3 there are 10 questions that can be used to evaluate an experimental design. Answer these 10 questions for the design proposed in Step 2.

4. After evaluating your proposed design, are there any changes you would like to make to your design? Explain.
When individuals climb to high altitudes, a condition known as acute mountain sickness (AMS) may occur. AMS is brought about by a combination of reduced air pressure and lower oxygen concentration that occurs at high altitudes. Two standard treatments for AMS are a medication, acetazolamide (which stimulates breathing and reduces mild symptoms) and the use of portable hyperbaric chambers.

With increasing numbers of younger inexperienced mountaineers, it is important to re-evaluate these treatments for the 12 to 14 year age group. An experimental plan under consideration is to study the first 18 youngsters diagnosed with AMS at a high altitude park ranger station whose parents consent to participation in the experiment. Equal numbers of each treatment are desired and the researchers are considering the following strategy for random assignment of treatments: Assign the treatments using a coin flip until one treatment has been assigned nine times; then assign the other treatment to the remaining subjects.

The table below presents data on the first 18 young climbers whose parents consented to participation in the experiment.

<table>
<thead>
<tr>
<th>Order</th>
<th>Gender</th>
<th>Age (yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>male</td>
<td>12.90</td>
</tr>
<tr>
<td>2</td>
<td>female</td>
<td>13.34</td>
</tr>
<tr>
<td>3</td>
<td>male</td>
<td>12.39</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>13.95</td>
</tr>
<tr>
<td>5</td>
<td>male</td>
<td>13.63</td>
</tr>
<tr>
<td>6</td>
<td>male</td>
<td>13.62</td>
</tr>
<tr>
<td>7</td>
<td>female</td>
<td>12.55</td>
</tr>
<tr>
<td>8</td>
<td>female</td>
<td>13.54</td>
</tr>
<tr>
<td>9</td>
<td>male</td>
<td>12.34</td>
</tr>
<tr>
<td>10</td>
<td>female</td>
<td>13.74</td>
</tr>
<tr>
<td>11</td>
<td>female</td>
<td>13.78</td>
</tr>
<tr>
<td>12</td>
<td>male</td>
<td>14.05</td>
</tr>
<tr>
<td>13</td>
<td>female</td>
<td>14.22</td>
</tr>
<tr>
<td>14</td>
<td>female</td>
<td>13.91</td>
</tr>
<tr>
<td>15</td>
<td>male</td>
<td>14.39</td>
</tr>
<tr>
<td>16</td>
<td>female</td>
<td>13.54</td>
</tr>
<tr>
<td>17</td>
<td>female</td>
<td>13.85</td>
</tr>
<tr>
<td>18</td>
<td>male</td>
<td>14.11</td>
</tr>
</tbody>
</table>

1. Describe how you would implement a strategy equivalent to the one proposed by the researchers. Your plan should assign the treatments M (medicine) and H (hyperbaric chamber) to these climbers as they appear at the ranger station.
2. Implement your strategy in Step (1), assigning treatments to climbers 1–18.
3. Looking at which climbers were assigned to each of the two groups, do you feel that this method worked well? Why or why not?
4. Compute the proportion of females in the medicine group. How does this proportion compare to the proportion of females in the entire group of 18 subjects?
5. Construct two dotplots—one of the ages of those assigned to the medicine treatment and one of the ages of those assigned to the hyperbaric chamber treatment. Are the age distributions for the two groups similar?
6. Compute the average age of those assigned to the medicine group. How does it compare to the average age for the other treatment group?
7. Record the proportion of females in the medicine group, the average age of those assigned to the medicine group, and the average age of those assigned to the hyperbaric chamber group obtained by each student on your class.
8. Using the values from Step (6), construct a dotplot of each of the following: the proportion of females in the medicine group, the average age of those assigned to the medicine group, and the average age of those assigned to the hyperbaric chamber group.
9. Using the results of the previous steps, evaluate the success of this random assignment strategy. Write a short paragraph explaining to the researchers whether or not they should use the proposed strategy for random assignment and why.
### Summary of Key Concepts and Formulas

<table>
<thead>
<tr>
<th>TERM OR FORMULA</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational study</td>
<td>A study that observes characteristics of an existing population.</td>
</tr>
<tr>
<td>Simple random sample</td>
<td>A sample selected in a way that gives every different sample of size $n$ an equal chance of being selected.</td>
</tr>
<tr>
<td>Stratified sampling</td>
<td>Dividing a population into subgroups (strata) and then taking a separate random sample from each stratum.</td>
</tr>
<tr>
<td>Cluster sampling</td>
<td>Dividing a population into subgroups (clusters) and forming a sample by randomly selecting clusters and including all individuals or objects in the selected cluster in the sample.</td>
</tr>
<tr>
<td>1 in $k$ systematic sampling</td>
<td>A sample selected from an ordered arrangement of a population by choosing a starting point at random from the first $k$ individuals on the list and then selecting every $k$th individual thereafter.</td>
</tr>
<tr>
<td>Confounding variable</td>
<td>A variable that is related both to group membership and to the response variable.</td>
</tr>
<tr>
<td>Measurement or response bias</td>
<td>The tendency for samples to differ from the population because the method of observation tends to produce values that differ from the true value.</td>
</tr>
<tr>
<td>Selection bias</td>
<td>The tendency for samples to differ from the population because of systematic exclusion of some part of the population.</td>
</tr>
<tr>
<td>Nonresponse bias</td>
<td>The tendency for samples to differ from the population because measurements are not obtained from all individuals selected for inclusion in the sample.</td>
</tr>
<tr>
<td>Experiment</td>
<td>A procedure for investigating the effect of experimental conditions (treatments) on a response variable.</td>
</tr>
<tr>
<td>Treatments</td>
<td>The experimental conditions imposed by the experimenter.</td>
</tr>
<tr>
<td>Extraneous variable</td>
<td>A variable that is not an explanatory variable in the study but is thought to affect the response variable.</td>
</tr>
<tr>
<td>Direct control</td>
<td>Holding extraneous variables constant so that their effects are not confounded with those of the experimental conditions.</td>
</tr>
<tr>
<td>Blocking</td>
<td>Using extraneous variables to create groups that are similar with respect to those variables and then assigning treatments at random within each block, thereby filtering out the effect of the blocking variables.</td>
</tr>
</tbody>
</table>
Collecting Data Sensibly

<table>
<thead>
<tr>
<th>TERM OR FORMULA</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random assignment</td>
<td>Assigning experimental units to treatments or treatments to trials at random.</td>
</tr>
<tr>
<td>Replication</td>
<td>A strategy for ensuring that there is an adequate number of observations on each experimental treatment.</td>
</tr>
<tr>
<td>Placebo treatment</td>
<td>A treatment that resembles the other treatments in an experiment in all apparent ways but that has no active ingredients.</td>
</tr>
<tr>
<td>Control group</td>
<td>A group that receives no treatment.</td>
</tr>
<tr>
<td>Single-blind experiment</td>
<td>An experiment in which the subjects do not know which treatment they received but the individuals measuring the response do know which treatment was received, or an experiment in which the subjects do know which treatment they received but the individuals measuring the response do not know which treatment was received.</td>
</tr>
<tr>
<td>Double-blind experiment</td>
<td>An experiment in which neither the subjects nor the individuals who measure the response know which treatment was received.</td>
</tr>
</tbody>
</table>

Chapter Review Exercises 2.70 - 2.85

2.70 A pollster for the Public Policy Institute of California explains how the Institute selects a sample of California adults ("It's About Quality, Not Quantity," San Luis Obispo Tribune, January 21, 2000):

That is done by using computer-generated random residential telephone numbers with all California prefixes, and when there are no answers, calling back repeatedly to the original numbers selected to avoid a bias against hard-to-reach people. Once a call is completed, a second random selection is made by asking for the adult in the household who had the most recent birthday. It is as important to randomize who you speak to in the household as it is to randomize the household you select. If you didn’t, you’d primarily get women and older people.

Comment on this approach to selecting a sample. How does the sampling procedure attempt to minimize certain types of bias? Are there sources of bias that may still be a concern?

2.71 Based on a survey of 4113 U.S. adults, researchers at Stanford University concluded that Internet use leads to increased social isolation. The survey was conducted by an Internet-based polling company that selected its samples from a pool of 35,000 potential respondents, all of whom had been given free Internet access and WebTV hardware in exchange for agreeing to regularly participate in surveys conducted by the polling company. Two criticisms of this study were expressed in an article that appeared in the San Luis Obispo Tribune (February 28, 2000). The first criticism was that increased social isolation was measured by asking respondents if they were talking less to family and friends on the phone. The second criticism was that the sample was selected only from a group that was induced to participate by the offer of free Internet service, yet the results were generalized to all U.S. adults. For each criticism, indicate what type of bias is being described and why it might make you question the conclusion drawn by the researchers.
The article "I'd Like to Buy a Vowel, Drivers Say" (USA Today, August 7, 2001) speculates that young people prefer automobile names that consist of just numbers and/or letters that do not form a word (such as Hyundai's XG300, Mazda's 626, and BMW's 325i). The article goes on to state that Hyundai had planned to identify the car that was eventually marketed as the XG300 with the name Concerto, until they determined that consumers hated it and that they thought XG300 sounded more "technical" and deserving of a higher price. Do the students at your school feel the same way? Describe how you would go about selecting a sample to answer this question.

A study in Florida is examining whether health literacy classes and using simple medical instructions that include pictures and avoid big words and technical terms can keep Medicaid patients healthier (San Luis Obispo Tribune, October 16, 2002). Twenty-seven community health centers are participating in the study. For 2 years, half of the centers will administer standard care. The other centers will have patients attend classes and will provide special health materials that are easy to understand. Explain why it is important for the researchers to assign the centers to the two groups (standard care and classes with simple health literature) at random.

Is status related to a student’s understanding of science? The article “From Here to Equity: The Influence of Status on Student Access to and Understanding of Science” (Culture and Comparative Studies [1999]: 577–602) described a study on the effect of group discussions on learning biology concepts. An analysis of the relationship between status and “rate of talk” (the number of on-task speech acts per minute) during group work included gender as a blocking variable. Do you think that gender is a useful blocking variable? Explain.

The article “Tots’ TV-Watching May Spur Attention Problems” (San Luis Obispo Tribune, April 5, 2004) describes a study that appeared in the journal Pediatrics. In this study, researchers looked at records of 2500 children who were participating in a long-term health study. They found that 10% of these children had attention disorders at age 7 and that hours of television watched at age 1 and age 3 was associated with an increased risk of having an attention disorder at age 7.

a. Is the study described an observational study or an experiment?

b. Give an example of a potentially confounding variable that would make it unwise to draw the conclusion that hours of television watched at a young age is the cause of the increased risk of attention disorder.

A study of more than 50,000 U.S. nurses found that those who drank just one soda or fruit punch a day tended to gain much more weight and had an 80% increased risk in developing diabetes compared to those who drank less than one a month. (The Washington Post, August 25, 2004). “The message is clear. . . . Anyone who cares about their health or the health of their family will not consume these beverages,” said Walter Willett of the Harvard School of Public Health, who helped conduct the study. The sugar and beverage industries said that the study was fundamentally flawed. “These allegations are inflammatory. Women who drink a lot of soda may simply have generally unhealthy lifestyles,” said Richard Adamson of the American Beverage Association.

a. Do you think that the study described was an observational study or an experiment?

b. Is it reasonable to conclude that drinking soda or fruit punch causes the observed increased risk of diabetes? Why or why not?

“Crime Finds the Never Married” is the conclusion drawn in an article from USA Today (June 2001). This conclusion is based on data from the Justice Department’s National Crime Victimization Survey which estimated the number of violent crimes per 1000 people, 12 years of age or older, to be 51 for the never married, 42 for the divorced or separated, 13 for married individuals, and 8 for the widowed. Does being single cause an increased risk of violent crime? Describe a potential confounding variable that illustrates why it is unreasonable to conclude that a change in marital status causes a change in crime risk.

The article “Workers Grow More Dissatisfied” in the San Luis Obispo Tribune (August 22, 2002) states that “a survey of 5000 people found that while most Americans continue to find their jobs interesting and are even satisfied with their commutes, a bare majority like their jobs.” This statement was based on the finding that only 51 percent of those responding to a mail survey indicated that they were satisfied with their jobs. Describe any potential sources of bias that might limit the researcher's ability to draw conclusions about working Americans based on the data collected in this survey.
According to the article “Effect of Preparation Methods on Total Fat Content, Moisture Content, and Sensory Characteristics of Breaded Chicken Nuggets and Beef Steak Fingers” (Family and Consumer Sciences Research Journal [1999]: 18–27), sensory tests were conducted using 40 college student volunteers at Texas Women’s University. Give three reasons, apart from the relatively small sample size, why this sample may not be ideal as the basis for generalizing to the population of all college students.

Do ethnic group and gender influence the type of care that a heart patient receives? The following passage is from the article “Heart Care Reflects Race and Sex, Not Symptoms” (USA Today, February 25, 1999, reprinted with permission):

Previous research suggested blacks and women were less likely than whites and men to get cardiac catheterization or coronary bypass surgery for chest pain or a heart attack. Scientists blamed differences in illness severity, insurance coverage, patient preference, and health care access. The researchers eliminated those differences by videotaping actors—two black men, two black women, two white men, and two white women—describing chest pain from identical scripts. They wore identical gowns, used identical gestures, and were taped from the same position. Researchers asked 720 primary care doctors at meetings of the American College of Physicians or the American Academy of Family Physicians to watch a tape and recommend care. The doctors thought the study focused on clinical decision making.

Evaluate this experimental design. Do you think this is a good design or a poor design, and why? If you were designing such a study, what, if anything, would you propose to do differently?

An article in the San Luis Obispo Tribune (September 7, 1999) described an experiment designed to investigate the effect of creatine supplements on the development of muscle fibers. The article states that the researchers “looked at 19 men, all about 25 years of age and similar in weight, lean body mass, and capacity to lift weights. Ten were given creatine—25 grams a day for the first week, followed by 5 grams a day for the rest of the study. The rest were given a fake preparation. No one was told what he was getting. All the men worked out under the guidance of the same trainer. The response variable measured was gain in fat-free mass (in percent).”

What extraneous variables are identified in the given statement, and what strategy did the researchers use to deal with them?

Do you think it was important that the men participating in the experiment were not told whether they were receiving creatine or the placebo? Explain.

This experiment was not conducted in a double-blind manner. Do you think it would have been a good idea to make this a double-blind experiment? Explain.

Researchers at the University of Houston decided to test the hypothesis that restaurant servers who squat to the level of their customers would receive a larger tip (“Effect of Server Posture on Restaurant Tipping,” Journal of Applied Social Psychology [1993]: 678–685). In the experiment, the waiter would flip a coin to determine whether he would stand or squat next to the table. The waiter would record the amount of the bill and of the tip and whether he stood or squatted.

Describe the treatments and the response variable.

Discuss possible extraneous variables and how they could be controlled.

Discuss whether blocking would be necessary.

Identify possible confounding variables.

Discuss the role of random assignment in this experiment.

You have been asked to determine on what types of grasslands two species of birds, northern harriers and short-eared owls, build nests. The types of grasslands to be used include undisturbed native grasses, managed native grasses, undisturbed nonnative grasses, and managed non-native grasses. You are allowed a plot of land 500 meters square to study. Explain how you would determine where to plant the four types of grasses. What role would random assignment play in this determination? Identify any confounding variables. Would this study be considered an observational study or an experiment? (Based on the article “Response of Northern Harriers and Short-Eared Owls to Grassland Management in Illinois,” Journal of Wildlife Management [1999]: 517–523.)

A manufacturer of clay roofing tiles would like to investigate the effect of clay type on the proportion of tiles that crack in the kiln during firing. Two different types of clay are to be considered. One hundred tiles can be placed in the kiln at any one time. Firing temperature varies slightly at different locations in the kiln, and firing temperature may also affect cracking. Discuss the design of an experiment to collect information that could be used to plant the four types of grasses. What role would random assignment play in this determination? Identify any confounding variables. Would this study be considered an observational study or an experiment? (Based on the article “Response of Northern Harriers and Short-Eared Owls to Grassland Management in Illinois,” Journal of Wildlife Management [1999]: 517–523.)
used to decide between the two clay types. How does your proposed design deal with the extraneous variable temperature?

2.85 A mortgage lender routinely places advertisements in a local newspaper. The advertisements are of three different types: one focusing on low interest rates, one featuring low fees for first-time buyers, and one appealing to people who may want to refinance their homes. The lender would like to determine which advertisement format is most successful in attracting customers to call for more information. Describe an experiment that would provide the information needed to make a determination. Be sure to consider extraneous variables, such as the day of the week that the advertisement appears in the paper, the section of the paper in which the advertisement appears, or daily fluctuations in the interest rate. What role does random assignment play in your design?

Graphing Calculator Explorations

EXPLORATION 2.1 Calculators and the Study of Statistics

You must be able to use your calculator in order to be able to analyze data. In previous math classes you may have used your calculator for graphing functions, finding solutions to equations, and arithmetic calculations. In statistics you will use your calculator differently and will also use new calculator keys and menu items. Graphing Calculator Explorations are intended to help you get maximum utility from your calculator. These explorations highlight some important features of your calculator. In order to speak to the widest possible audience, the explorations will be generic in nature, rather than showcasing a particular calculator.

Calculators vary in statistical capability and in the applications that can be downloaded from the web. The characteristics of a graphing calculator that are important for the study of statistics are:

• Capability to perform elementary statistical calculations (computing means, standard deviations, correlation coefficients, and regression equations)
• Capability to generate statistical graphs (boxplots, histograms, and scatterplots)
• Row-and-column data entry format

EXPLORATION 2.2 Generating Random Integers

Procedures for generating random numbers have been around for a long time. The earliest techniques for generating random numbers were throwing dice, dealing cards, and selecting well-mixed numbered balls from a container. Today, computers and calculators use algorithms to generate “random” numbers. These algorithms generate what are called “pseudo-random” numbers. For most purposes, the “random” numbers generated by today’s computers and calculators are perfectly fine, and we refer to these numbers as random.

Generating random numbers is a built-in function of graphing calculators. To learn the appropriate keystrokes for your calculator you will need to consult your calculator manual; look for “rand,” or possibly “random” in the index. On some calculators the precision can be adjusted and in our discussions we will generally use four digit accuracy. On most calculators, a single keystroke or a short sequence of keystrokes will produce a random number between 0 and 1. Locate that function and then generate five random numbers using the procedures presented in your manual. Some calculators will repeat the process each time you press “enter” or “execute.” Pressing the enter/execute button four times after you get the first random number. If a random number appears each time, smile—this will save many keystrokes!
numbers we obtained (yours will be different) are 0.5147, 0.4058, 0.7338, 0.0440, and 0.3394. Random number generators typically produce a random number, \( r \), such that \( 0 \leq r < 1 \). Your calculator may have additional random number functions, such as \text{randInt()} \) to generate integers. To see if your calculator has this built-in capability, look in the manual’s index for something like “random integer.” In the discussions that follow, we will \textit{not} assume you have a random integer capability, but if you do, please use it!

If you do not have a built-in capability to generate random integers, the method in the box below will be helpful.

---

**Converting Random Numbers to Integers, 1, 2, 3, \ldots, n**

To convert a calculator-generated decimal random number \( r \), \( 0 \leq r < 1 \), into a random integer in the range 1, 2, 3, \ldots, \( n \), multiply \( r \) by \( n \), add 1, and ignore the digits to the right of the decimal.

This will involve a sequence of keystrokes something like the following, where we use “\textit{rand}” to mean the keystrokes needed to generate \( r \):

\[
\text{rand} \times n + 1.
\]

Most scientific calculators have an “\textit{int}” or “\textit{floor}” function, which will truncate the decimal by “rounding down.” If your calculator has this capability, you can accomplish the integer random number generation in one sequence of keystrokes:

\[
\text{int(}\text{rand} \times n\text{)} + 1.
\]

---

**Example: Generating Random Numbers**

Generate five random integers between 1 and 100 for purposes of sampling from a population with 100 individuals. The keystrokes to generate a random integer between 1 and 100 are:

\[
\text{int(}\text{rand} \times 100\text{)} + 1.
\]

(Remember, \textit{rand} stands for the sequence of keystrokes needed to get the random number between 0 and 1 and \textit{int} stands for the keystrokes necessary to truncate a decimal to an integer.)

The numbers we obtained (though of course your numbers will differ) are

11  55  29  38  37

We would then include the individuals identified by these numbers in the sample. (Note that in this example we are sampling from a population. If we were sampling without replacement and the random number generated by the calculator resulted in two or more random integers that were the same, we would have ignored the duplicates and generated additional random integers as needed.)

Adding 1 in the formula above is not some sort of magic. Because the \textit{rand} keystrokes return random numbers in the range \( 0 \leq r < 1 \), it is possible for \textit{rand} to deliver a 0. If you are not bothered by random integers starting at 0, you need not waste keystrokes by adding 1 each time. Also note that the arithmetic random number generators in calculators are shipped from the factory with a number called a “seed,” needed to start the random number generation process. If two calculators in your class are “right out of the box” you may notice that the random numbers generated by these calculators are the same. This will not be a problem for very long; calculator users will typically press the \textit{rand} sequence different numbers of times, and sequences will soon differ in actual classroom use.
EXPLORATION 2.3  Random Assignment to Treatments

The process of random assignment of subjects to treatments is critical in the proper design of an experiment. Random assignment can be accomplished using the graphing calculator’s capability to generate integers from 1, 2 \ldots n. We will illustrate how this can be done in some common experimental situations.

In the first experiment, we will assign subjects to treatments without trying to get the same number of subjects in each treatment group. This experiment investigates the effect of pizza on performance on a statistics exam. An instructor has decided to use three types of pizza (the treatments): sausage pizza, mushroom pizza, and cheese pizza. Twelve randomly selected students will take part in the experiment. The instructor’s strategy for assigning treatments is very simple: Generate 12 random integers from the list \{1, 2, 3\}. Before any students are assigned, he arbitrarily assigns mushroom pizza = 1, cheese pizza = 2, and sausage pizza = 3.

We now randomly generate integers between 1 and 3 using

\[ \text{int}(\text{rand} \times 3) + 1. \]

Our results were (remember that yours will surely be different)

3, 1, 1, 1, 1, 3, 1, 2, 1, 3, 1, 3

These numbers are used to assign each student to a treatment (pizza type) in the experiment, as shown. Entries in the table are student, treatment number, and (treatment).

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>(Sausage)</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>(Mushroom)</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>(Mushroom)</td>
<td>10</td>
</tr>
</tbody>
</table>

We now randomly generate integers between 1 and 3 using

\[ \text{int}(\text{rand} \times 3) + 1. \]

These numbers are used to assign each student to a treatment (pizza type) in the experiment, as shown. Entries in the table are student, treatment number, and (treatment).

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>(Sausage)</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>(Mushroom)</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>(Mushroom)</td>
<td>10</td>
</tr>
</tbody>
</table>

We can quickly see some problems using this method of assignment: More than 50% of the subjects were assigned to mushroom pizza, and there is no replication of the cheese pizza treatment! What if we want to have equal-sized treatment groups? We might try to achieve equal treatment group sizes by adding a rule to our procedure: “Once any treatment group has four subjects assigned, do not assign any more subjects to that treatment.” Unfortunately, this approach can cause problems (see discussion on random assignment in Section 2.3). Fortunately, we have another strategy.

First, we number the subjects from 1–12. Then we generate 12 random numbers between 0 and 1, ignoring any repeats. The first four random numbers generated are associated with treatment 1, the next four with treatment 2, and finally the last four random numbers are associated with treatment 3. One possible result of this process is shown below.

\[ .5441 .6379 .9295 .6742 .7980 .5522 .3377 .2187 .6217 .8811 .1243 .4500 \]

Finally, subject 1 is associated with the smallest of the 12 random numbers; subject 2 is associated with the second smallest random number, and so on. This would result in the following assignment of subjects to treatments.

\[ .5441 .6379 .9295 .6742 .7980 .5522 .3377 .2187 .6217 .8811 .1243 .4500 \]

Copyright 2011 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.
In this random assignment of subjects to treatments, subjects 5, 8, 12, and 9 get mushroom pizza; subjects 10, 6, 3, and 2 get cheese pizza; and the rest get sausage pizza.

**Random Assignment to Treatments—with Blocking**  Random assignment in a situation in which the experimenter is using blocking to control an extraneous variable could also be implemented using random integers. Suppose the instructor in the previous example has some seats near the window and some seats that are not by the window. It is possible that students near the window might be distracted and that this might affect exam performance. Because of this, it would be reasonable to block by position in the room. Suppose that six students will participate in the experiment. We need to consider the blocking strategy as we assign treatments.

Our random assignment is restricted a bit compared to what we did earlier because now we need to have each treatment represented in each block. Suppose that the six students are seated for the exam in the arrangement shown in the table below.

<table>
<thead>
<tr>
<th>Near Window</th>
<th>Not Near Window</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We now want to assign treatments at random to the students within each block. A sequence of random integers between 1 and 3 can be used for this purpose. One possible sequence is:

\[3 \ 3 \ 2 \ 1 \ 1 \ 3 \ 1 \ 2 \ 1 \ 1 \ 3 \ 2\]

The treatments are assigned as shown in the table below, starting with the “Near Window” block and then, once the assignment for the “Near Window” block is completed, moving to the “Not Near Window” block. Remember that mushroom pizza = 1, cheese pizza = 2, and sausage pizza = 3.
The final assignment of treatments to subjects within each block is shown in the table below. Note that each treatment appears once in each block.

<table>
<thead>
<tr>
<th>Random Integer</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Assign student in row one of near window block to sausage pizza treatment</td>
</tr>
<tr>
<td>3</td>
<td>Ignore, as there is already a student assigned to sausage pizza in the near window block</td>
</tr>
<tr>
<td>2</td>
<td>Assign student in row two of near window block to cheese pizza treatment</td>
</tr>
<tr>
<td>—</td>
<td>Assign student in row three of near window block to mushroom pizza treatment (since that is the only treatment that has not already been assigned in this block)</td>
</tr>
<tr>
<td>1</td>
<td>Assign student in row one of not near window block to mushroom pizza treatment</td>
</tr>
<tr>
<td>1</td>
<td>Ignore, as there is already a student assigned to mushroom pizza in the near window block</td>
</tr>
<tr>
<td>3</td>
<td>Assign student in row two of not near window block to sausage pizza treatment</td>
</tr>
<tr>
<td>—</td>
<td>Assign student in row three of near window block to cheese pizza treatment (since that is the only treatment that has not already been assigned in this block)</td>
</tr>
</tbody>
</table>

The need for random real numbers sometimes arises in sampling situations in which it is not possible to make a list of the elements of the population. For example, suppose we want to study chemical residue from the use of pesticides on farms in Iowa. A map of Iowa is overlaid with a grid of 1-mile by 1-mile squares. Picking squares at random in order to measure chemical residue would be a good start, but within any randomly selected square the chemical residue measurement must be taken at some actual location in the square. To determine this location, imagine a grid overlaying the map of Iowa, and using a random real number to select a point within each square.

**EXPLORATION 2.4 Generating Random Real Numbers**

In previous Explorations, we used the `rand` function to generate random numbers between 0 and 1 and then converted them to integers. In this Exploration, we use the `rand` function for generating random real numbers in some interval $(a, b)$.

Now it’s your turn to try it.

Acupuncture is a popular alternative treatment for chronic back pain. Does the treatment work, or is it psychological? To try to find out, an experiment is planned at a local clinic. As the statistical expert of the day, you have been asked to assign treatments (real or simulated acupuncture). Use the method discussed in this exploration to assign 12 subjects to the two treatments.
coordinate system with the origin in the southwest corner of a square of land. Since each square is 1 mile by 1 mile, generating a random point in a square of land is easy using the \texttt{rand} function: generate two random real numbers between 0 and 1, and use them as \(x\)- and \(y\)-coordinates for the square. Suppose the \texttt{rand} function generated 0.45059 and 0.98906. We would convert this to (0.45059, 0.98906), which would then determine the location where the measurement should be made.

Suppose now that we want to select random locations from a map of the wilds of Saskatchewan, Canada, in order to estimate the amount of a particular mineral in the soil. Here we can use the “natural” coordinates provided by the latitude and longitude of the map to determine location. Saskatchewan’s latitudes range from 49 to 60N, and longitudes from 102 to 110W. For generating random points in Saskatchewan, we can use the \texttt{rand} function and a little algebra.

We approach this problem in stages. First, consider the problem of generating a random number between 0 and \(b\). One solution is to generate a random number between 0 and 1, and multiply it by the positive real number, \(b\). Algebraically we see that if \(\texttt{rand}\) represents the random number generated,

\[
0 \leq \texttt{rand} < 1 \\
\texttt{rand} (0) \leq b \times (\texttt{rand}) < b \times (1) \\
0 \leq b \times (\texttt{rand}) < b
\]

Our ability to generate random real numbers is not limited to intervals that begin at 0. To generate a random real number between any two real numbers, \(a\) and \(b\), we can do a bit more algebra:

\[
0 \leq \texttt{rand} < 1 \\
(b - a) \times (0) \leq (b - a) \times (\texttt{rand}) < (b - a) \times (1) \\
0 \leq (b - a) \times (\texttt{rand}) < (b - a) \\
0 + a \leq (b - a) \times (\texttt{rand}) + a < (b - a) + a \\
a \leq (b - a) \times (\texttt{rand}) + a < b
\]

Saskatchewan is located between the latitudes 49 and 60N and longitudes 102 and 110W. So,

\[
102 \leq (110 - 102) \times (\texttt{rand}) + 102 < 110
\]
generates a random longitude in Saskatchewan and

\[
49 \leq (60 - 49) \times (\texttt{rand}) + 49 < 60
\]
generates a random latitude in Saskatchewan.

For example, using \texttt{rand}, we might get: 0.6217 and 0.8811. Substituting these values into the formulas above results in this random location:

\[
(110 - 102) \times (0.6217) + 102 = 106.97^\circ W \\
(60 - 49) \times (0.8811) + 49 = 58.69^\circ N
\]

This random location is somewhere close to Fond du Lac, on Lake Athabasca, just down the road from Uranium City!

\textbf{Now it’s your turn to try it.}

To provide some quick practice with generating random numbers using your calculator and also to learn about an interesting use of random numbers, we will find an area using what are known as “Monte Carlo” methods. We will learn something of the mathematics and statistics behind why this works in later chapters, but you can watch it unfold now.
Our story begins with a simple 1-inch by 1-inch square, and a diagonal line segment as shown below. We will use random numbers to approximate the area inside the square and below the line. Of course, as we remember from geometry, the area will be 0.5 square inches. The really neat thing is that we can use this method when the line turns into any mathematical function.

If we imagine a coordinate system with the origin at the lower left corner of the square, the line is: $y = x$. The Monte Carlo method of calculating an area hinges on this fact: If we generate random points inside the square, the proportion of the points “below the line” will reflect the area under the line. (We made the area of the square equal one for easy calculation.) The more points we generate, the closer the proportion of points below the line will be to—in this case—0.5.

To see this unfold, generate 20 random numbers and fill in the table below. (It doesn’t matter whether you go across or down as you fill the table.)

Now consider the ordered pairs $(x, y)$ from the columns of the table. If $x$ is less than $y$, the point will be above the line. If $x$ is greater than $y$, the point will be below the line. Approximately 5 of these points should be below the line! If you think your results are just lucky random numbers, try another 10 points—how many of the 20 points are below the line? If you are really suspicious, try another 20. If you don’t believe, keep adding points. Keep track of the proportion of points below the line after each additional 5 points—the proportion gets closer and closer to 0.5 as the number of points generated gets larger.
1. College students who reported being Internet users were surveyed, and data on the following attributes were recorded:

I. number of times they accessed the Internet in the previous 24 hours
II. length of the most recent Internet session
III. purpose of the last Internet session (e-mail, shopping, downloading music, etc.)
IV. approximate number of e-mails sent per day

Which of these variables is/are numerical?

(A) II only
(B) III only
(C) I and IV only
(D) I, II, and IV only
(E) I, III, and IV only

2. Which of the following numerical variable(s) are discrete?

I. number of books in a student’s backpack
II. weight of a student’s backpack
III. number of students in a class who carry a backpack
IV. thickness of the fabric used in making a student’s backpack

(A) I only
(B) II only
(C) I and III only
(D) I and IV only
(E) I, II, III, and IV are all discrete

3. The term used to describe the bias that occurs if some segment of a population is systematically excluded from a sample is

(A) selection bias.
(B) measurement bias.
(C) response bias.
(D) exclusion bias.
(E) visibility bias.

4. Two variables are confounded if

(A) their effects on the treatments cannot be ascertained.
(B) their effects on the sampling design cannot be observed.

(C) their effects on the response variable cannot be distinguished.
(D) their effects on the experimental design cannot be justified.
(E) their effects on the blocking factor cannot be measured.

5. In utilizing direct control, which of the following are held constant?

(A) values of an extraneous variable
(B) values of a blocking variable
(C) values of a response variable
(D) values of an explanatory variable
(E) values of a lurking variable

6. Fifty college students were asked what type of movie they had seen most recently. The responses are summarized in the frequency distribution shown.

<table>
<thead>
<tr>
<th>Type of Movie</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comedy</td>
<td>23</td>
</tr>
<tr>
<td>Drama</td>
<td>9</td>
</tr>
<tr>
<td>Action</td>
<td>7</td>
</tr>
<tr>
<td>Western</td>
<td>4</td>
</tr>
<tr>
<td>Animated</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

What is the relative frequency for the comedy category?

(A) 23
(B) 23/27
(C) 0.23
(D) 0.46
(E) 46

7. In a study to determine if the color of the label might affect whether potential customers would buy a particular brand of bottled tea, 100 volunteer subjects were recruited. They were shown two different label designs and asked which design they would be most likely to purchase. One label was blue with a photograph of a woman drinking the tea. The other label was green and did not include the photo. The volunteers were divided at random into two groups. One group was shown the blue label first, followed by the green label. The other group was shown the green label first, followed by the blue label.

*AP and the Advanced Placement Program are registered trademarks of the College Entrance Examination Board, which was not involved in the production and does not endorse, this product.
label first followed by the blue label. Which of the following is confounded with the treatments?

(A) color of cover (blue or green)
(B) cover composition (photo or no photo)
(C) preferred label (which label the person was most likely to purchase)
(D) group membership (blue first or green first)
(E) There are no confounding variables.

8. In order to estimate the proportion of students at a college who spend more than 2 hours per day on Facebook, a random sample of students at the college is selected and each student is interviewed about his or her use of Facebook. The students conducting the survey are worried that people who spend a lot of time on Facebook might be embarrassed to admit it and that their responses to the survey might not be honest. What type of bias are the students conducting the survey worried about?

(A) selection bias
(B) nonresponse bias
(C) measurement or response bias
(D) bias due to confounding
(E) They shouldn’t worry—there is no obvious source of bias.

9. To estimate the proportion of students who plan to purchase tickets to an upcoming school fundraiser, a high school decides to sample 100 students as they register for the spring semester. There are 2000 students at the school. Which of the following sampling plans would result in a stratified random sample?

(A) Number the students from 1 to 2000 and then use random numbers to select 100 students.
(B) Survey the first 100 students to register.
(C) Randomly select 100 students from a list of the 950 female students at the school.
(D) Divide the students into early registrants (the first 1000 to register) and late registrants (the last 1000 to register). Use random numbers to identify 50 of the early registrants and 50 of the late registrants to survey.
(E) Select one of the first 20 students to register using a random number table and then select every 20th student to register thereafter.

10. Which of the following describes a situation in which it is reasonable to reach a cause-and-effect conclusion based on data from a statistical study?

(A) The study is based on a random sample from a population of interest.
(B) The study is observational, and the sample used is not a convenience sample.
(C) The study is an experiment that uses random assignment to assign volunteers to experimental conditions (treatments).
(D) The study is observational, and the two samples used are not convenience samples.
(E) It is always reasonable to reach a cause-and-effect conclusion based on data from a statistical study.

11. A reporter for a local newspaper wants to survey county residents about their opinions on a proposal to raise property taxes to benefit the county library. He decides to ask 30 county residents whether they support this tax increase. He will select his sample by asking every third person entering the library starting at noon on a Friday. He will continue until he has asked a total of 30 county residents. Which type of bias is likely to be introduced by the way the sample will be selected?

(A) nonresponse bias
(B) response bias
(C) selection bias
(D) measurement bias
(E) No bias is introduced.

12. A random sample of 100 students at a particular college is to be selected, and each person selected will be asked how many times they went to a movie in a theater during the last year. The sample mean will then be used as an estimate of the mean number of times students at this college went to a movie in the last year. Which of the following is a reason to consider increasing the sample size for this study from 100 to 200?

(A) A larger sample size will reduce nonresponse bias.
(B) A larger sample size will reduce the response bias due to people not being able to accurately remember how many times they went to a movie in the last year.
(C) A larger sample size will reduce sampling variability—the differences in the sample mean that occur from sample to sample due to chance.
(D) It is less likely that one of the high values in the population (corresponding to a person who goes to a very large number of movies) will be included in a larger sample.
(E) All of the above are valid reasons to increase the sample size.
13. Researchers were interested in comparing students at private universities and students at public universities with respect to how much time they spent working in a typical week. They e-mailed a survey to 1000 students enrolled at a particular private university and to 2000 students enrolled at a particular public university. Data from the 400 private university students who responded to the survey and the 900 public university students who responded to the survey were used to determine that there was a significant difference in the mean number of hours worked in a typical week. Which of the following does not limit the researchers’ ability to generalize to the two populations of interest?

(A) The two sample sizes are not equal.
(B) There may be bias introduced due to nonresponse.
(C) The samples were not randomly selected.
(D) All students in the private university sample attended the same university, and all students in the public university sample attended the same university.
(E) All of the above limit the researchers’ ability to generalize to the populations of interest.

Use the following to answer questions 14–15.

One hundred volunteer subjects participated in a study to determine if room temperature affects people’s ability to concentrate. Female volunteers were given 10 minutes to try to memorize the words on a list of 50 nonsense words. The room temperature was controlled at 65 degrees (a cold room) while they completed the task. Male volunteers were also given 10 minutes to try to memorize the same list of words, but for the males, room temperature was controlled at 85 degrees (a hot room). At the end of the 10 minutes, each subject was asked to list as many of the words as he or she could remember, and the number correct was recorded. The resulting data were then used to determine if the mean number of words differed for the cold room and hot room conditions.

14. Which of the following is a confounding variable, that is, which of the following is confounded with the treatments?

(A) room temperature (cold or hot)
(B) gender (male or female)
(C) number of nonsense words recalled
(D) length of time given to memorize the words
(E) can’t tell because volunteers were used

15. The poor design of this experiment results in a variable that is confounded with the treatments. Which of the following changes to the design would be effective in eliminating this confounding?

(A) Use only one room temperature.
(B) Use only male subjects in the study and assign the males to one of the two room-temperature conditions at random.
(C) Assign the volunteers at random to one of the two room-temperature conditions.
(D) Create two blocks by putting all of the females in one block and all of the males in the other block. Then, within each block, assign subjects at random to one of the room-temperature conditions.
(E) B, C, and D are all strategies that would be effective.