11.2 Lesson

Core Vocabulary

population, *p. 604*
sample, *p. 604*
parameter, *p. 605*
statistic, *p. 605*
hypothesis, *p. 605*

Previous
Venn diagram
proportion

What You Will Learn

- Distinguish between populations and samples.
- Analyze hypotheses.

Populations and Samples

A **population** is the collection of all data, such as responses, measurements, or counts, that you want information about. A **sample** is a subset of a population.

A *census* consists of data from an entire population. But, unless a population is small, it is usually impractical to obtain all the population data. In most studies, information must be obtained from a *random sample*. (You will learn more about random sampling and data collection in the next section.)

It is important for a sample to be representative of a population so that sample data can be used to draw conclusions about the population. When the sample is not representative of the population, the conclusions may not be valid. Drawing conclusions about populations is an important use of *statistics*. Recall that statistics is the science of collecting, organizing, and interpreting data.

EXAMPLE 1 Distinguishing Between Populations and Samples

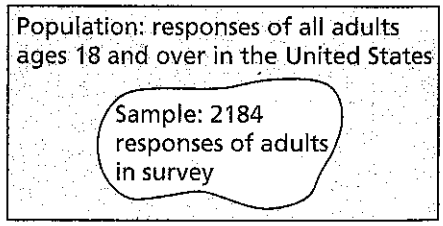
Identify the population and the sample. Describe the sample.

a. In the United States, a survey of 2184 adults ages 18 and over found that 1328 of them own at least one pet.

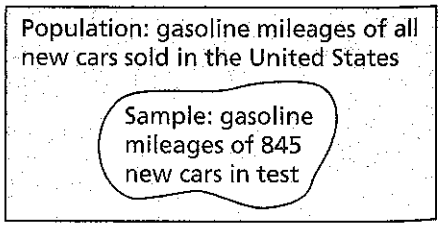
b. To estimate the gasoline mileage of new cars sold in the United States, a consumer advocacy group tests 845 new cars and finds they have an average of 25.1 miles per gallon.

SOLUTION

a. The population consists of the responses of all adults ages 18 and over in the United States, and the sample consists of the responses of the 2184 adults in the survey. Notice in the diagram that the sample is a subset of the responses of all adults in the United States. The sample consists of 1328 adults who said they own at least one pet and 856 adults who said they do not own any pets.

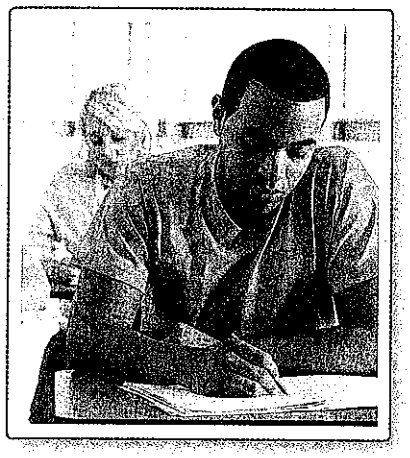


b. The population consists of the gasoline mileages of all new cars sold in the United States, and the sample consists of the gasoline mileages of the 845 new cars tested by the group. Notice in the diagram that the sample is a subset of the gasoline mileages of all new cars in the United States. The sample consists of 845 new cars with an average of 25.1 miles per gallon.



A numerical description of a population characteristic is called a **parameter**. A numerical description of a sample characteristic is called a **statistic**. Because some populations are too large to measure, a statistic, such as the sample mean, is used to estimate the parameter, such as the population mean. It is important that you are able to distinguish between a parameter and a statistic.

EXAMPLE 2 Distinguishing Between Parameters and Statistics



a. For all students taking the SAT in a recent year, the mean mathematics score was 514. Is the mean score a parameter or a statistic? Explain your reasoning.

b. A survey of 1060 women, ages 20–29 in the United States, found that the standard deviation of their heights is about 2.6 inches. Is the standard deviation of the heights a parameter or a statistic? Explain your reasoning.

SOLUTION

a. Because the mean score of 514 is based on all students who took the SAT in a recent year, it is a parameter.

b. Because there are more than 1060 women ages 20–29 in the United States, the survey is based on a subset of the population (all women ages 20–29 in the United States). So, the standard deviation of the heights is a statistic. Note that if the sample is representative of the population, then you can estimate that the standard deviation of the heights of all women ages 20–29 in the United States is about 2.6 inches.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

In Monitoring Progress Questions 1 and 2, identify the population and the sample.

1. To estimate the retail prices for three grades of gasoline sold in the United States, the Energy Information Association calls 800 retail gasoline outlets, records the prices, and then determines the average price for each grade.
2. A survey of 4464 shoppers in the United States found that they spent an average of $407.02 from Thursday through Sunday during a recent Thanksgiving holiday.
3. A survey found that the median salary of 1068 statisticians is about $72,800. Is the median salary a parameter or a statistic? Explain your reasoning.
4. The mean age of U.S. representatives at the start of the 113th Congress was about 57 years. Is the mean age a parameter or a statistic? Explain your reasoning.

UNDERSTANDING MATHEMATICAL TERMS

A *population proportion* is the ratio of members of a population with a particular characteristic to the total members of the population. A *sample proportion* is the ratio of members of a sample of the population with a particular characteristic to the total members of the sample.

Analyzing Hypotheses

In statistics, a **hypothesis** is a claim about a characteristic of a population. Here are some examples.

1. A drug company claims that patients using its weight-loss drug lose an average of 24 pounds in the first 3 months.
2. A medical researcher claims that the proportion of U.S. adults living with one or more chronic conditions, such as high blood pressure, is 0.45, or 45%.

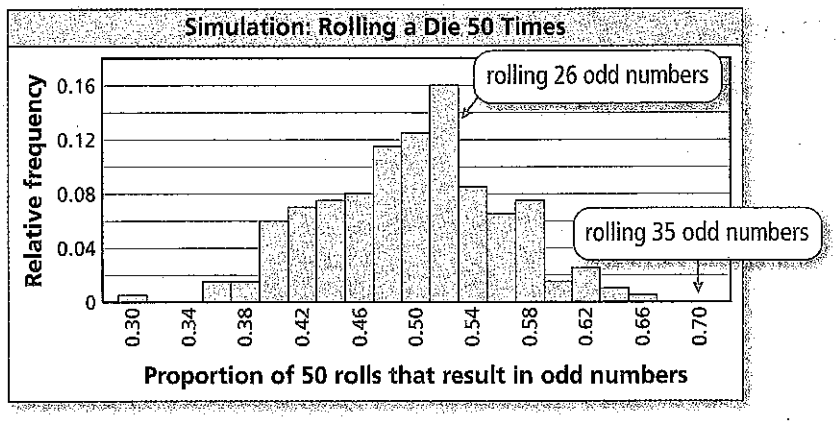
To analyze a hypothesis, you need to distinguish between results that can easily occur by chance and results that are highly unlikely to occur by chance. One way to analyze a hypothesis is to perform a *simulation*. When the results are highly unlikely to occur, the hypothesis is probably false.

EXAMPLE 3 Analyzing a Hypothesis

You roll a six-sided die 5 times and do not get an even number. The probability of this happening is $\left(\frac{1}{2}\right)^5 = 0.03125$, so you suspect this die favors odd numbers. The die maker claims the die does not favor odd numbers. What should you conclude when you roll the actual die 50 times and get (a) 26 odd numbers and (b) 35 odd numbers?

SOLUTION

The maker's claim, or hypothesis, is "the die does not favor odd numbers." This is the same as saying that the proportion of odd numbers rolled, in the long run, is 0.50. So, assume the probability of rolling an odd number is 0.50. Simulate the rolling of the die by repeatedly drawing 200 random samples of size 50 from a population of 50% ones and 50% zeros. Let the population of ones represent the event of rolling an odd number and make a histogram of the distribution of the sample proportions.



a. Getting 26 odd numbers in 50 rolls corresponds to a proportion of $\frac{26}{50} = 0.52$. In the simulation, this result had a relative frequency of 0.16. In fact, most of the results are close to 0.50. Because this result can easily occur by chance, you can conclude that the maker's claim is most likely true.

b. Getting 35 odd numbers in 50 rolls corresponds to a proportion of $\frac{35}{50} = 0.70$. In the simulation, this result did not occur. Because getting 35 odd numbers is highly unlikely to occur by chance, you can conclude that the maker's claim is most likely false.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

5. WHAT IF? In Example 3, what should you conclude when you roll the actual die 50 times and get (a) 24 odd numbers and (b) 31 odd numbers?

INTERPRETING MATHEMATICAL RESULTS

Results of other simulations may have histograms different from the one shown, but the shape should be similar. Note that the histogram is fairly bell-shaped and symmetric, which means the distribution is approximately normal. By increasing the number of samples or the sample sizes (or both), you should get a histogram that more closely resembles a normal distribution.

JUSTIFYING CONCLUSIONS

In Example 3(b), the theoretical probability of getting 35 odd numbers in 50 rolls is about 0.002. So, while unlikely, it is possible that you incorrectly concluded that the die maker's claim is false.

In Example 3(b), you concluded the maker's claim is probably false. In general, such conclusions may or may not be correct. For instance, you may conclude that a hypothesis is false when it is actually true. The table summarizes the incorrect and correct decisions that can be made about a hypothesis.

		Truth of Hypothesis	
		Hypothesis is true.	**Hypothesis is false.**
Decision	**You decide that the hypothesis is true.**	correct decision	incorrect decision
	You decide that the hypothesis is false.	incorrect decision	correct decision

11.3 Lesson

What You Will Learn

- Identify types of sampling methods in statistical studies.
- Recognize bias in sampling.
- Analyze methods of collecting data.
- Recognize bias in survey questions.

Core Vocabulary

random sample, *p. 610*
self-selected sample, *p. 610*
systematic sample, *p. 610*
stratified sample, *p. 610*
cluster sample, *p. 610*
convenience sample, *p. 610*
bias, *p. 611*
unbiased sample, *p. 611*
biased sample, *p. 611*
experiment, *p. 612*
observational study, *p. 612*
survey, *p. 612*
simulation, *p. 612*
biased question, *p. 613*

Previous
population
sample

Identifying Sampling Methods in Statistical Studies

The steps in a typical statistical study are shown below.

Identify the variable of interest and the population of the study. ⇨ Choose a sample that is representative of the population. ⇨ Collect data. ⇨ Organize and describe the data using a statistic. ⇨ Interpret the data, make inferences, and draw conclusions about the population.

There are many different ways of sampling a population, but a *random sample* is preferred because it is most likely to be representative of a population. In a **random sample**, each member of a population has an equal chance of being selected.

The other types of samples given below are defined by the methods used to select members. Each sampling method has its advantages and disadvantages.

Core Concept

Types of Samples

For a **self-selected sample**, members of a population can volunteer to be in the sample.

For a **systematic sample**, a rule is used to select members of a population. For instance, selecting every other person.

STUDY TIP

A stratified sample ensures that every segment of a population is represented.

For a **stratified sample**, a population is divided into smaller groups that share a similar characteristic. A sample is then randomly selected from each group.

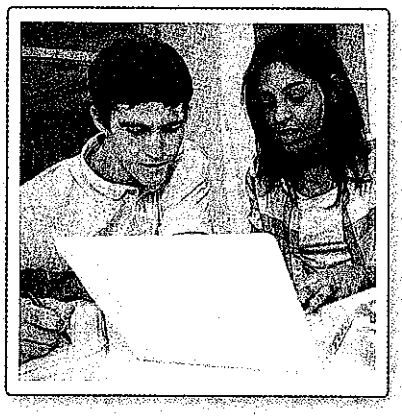
STUDY TIP

With cluster sampling, a member of a population cannot belong to more than one cluster.

For a **cluster sample**, a population is divided into groups, called *clusters*. All of the members in one or more of the clusters are selected.

For a **convenience sample**, only members of a population who are easy to reach are selected.

EXAMPLE 1 Identifying Types of Samples



You want to determine whether students in your school like the new design of the school's website. Identify the type of sample described.

a. You list all of the students alphabetically and choose every sixth student.

b. You mail questionnaires and use only the questionnaires that are returned.

c. You ask all of the students in your algebra class.

d. You randomly select two students from each classroom.

SOLUTION

a. You are using a rule to select students. So, the sample is a *systematic* sample.

b. The students can choose whether to respond. So, the sample is a *self-selected* sample.

c. You are selecting students who are readily available. So, the sample is a *convenience* sample.

d. The students are divided into similar groups by their classrooms, and two students are selected at random from each group. So, the sample is a *stratified* sample.

Monitoring Progress Help in English and Spanish at *BigIdeasMath.com*

1. **WHAT IF?** In Example 1, you divide the students in your school according to their zip codes, then select all of the students that live in one zip code. What type of sample are you using?
2. Describe another method you can use to obtain a stratified sample in Example 1.

Recognizing Bias in Sampling

STUDY TIP

All good sampling methods rely on random sampling.

A **bias** is an error that results in a misrepresentation of a population. In order to obtain reliable information and draw accurate conclusions about a population, it is important to select an *unbiased sample*. An **unbiased sample** is representative of the population that you want information about. A sample that overrepresents or under-represents part of the population is a **biased sample**. When a sample is biased, the data are invalid. A *random sample* can help reduce the possibility of a biased sample.

EXAMPLE 2 Identifying Bias in Samples

Identify the type of sample and explain why the sample is biased.

a. A news organization asks its viewers to participate in an online poll about bullying.

b. A computer science teacher wants to know how students at a school most often access the Internet. The teacher asks students in one of the computer science classes.

SOLUTION

a. The viewers can choose whether to participate in the poll. So, the sample is a *self-selected* sample. The sample is biased because people who go online and respond to the poll most likely have a strong opinion on the subject of bullying.

b. The teacher selects students who are readily available. So, the sample is a *convenience* sample. The sample is biased because other students in the school do not have an opportunity to be chosen.

EXAMPLE 3 Selecting an Unbiased Sample

You are a member of your school's yearbook committee. You want to poll members of the senior class to find out what the theme of the yearbook should be. There are 246 students in the senior class. Describe a method for selecting a random sample of 50 seniors to poll.

STUDY TIP

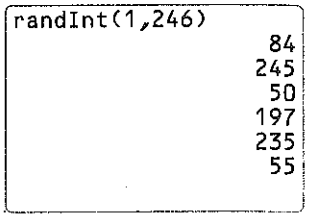
When you obtain a duplicate integer during the generation, ignore it and generate a new, unique integer as a replacement.

SOLUTION

Step 1 Make a list of all 246 seniors. Assign each senior a different integer from 1 to 246.

Step 2 Generate 50 unique random integers from 1 to 246 using the *randInt* feature of a graphing calculator.

Step 3 Choose the 50 students who correspond to the 50 integers you generated in Step 2.



Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

3. The manager of a concert hall wants to know how often people in the community attend concerts. The manager asks 45 people standing in line for a rock concert how many concerts they attend per year. Identify the type of sample the manager is using and explain why the sample is biased.

4. In Example 3, what is another method you can use to generate a random sample of 50 students? Explain why your sampling method is random.

Analyzing Methods of Data Collection

There are several ways to collect data for a statistical study. The objective of the study often dictates the best method for collecting the data.

Core Concept

Methods of Collecting Data

READING

A *census* is a survey that obtains data from every member of a population. Often, a census is not practical because of its cost or the time required to gather the data. The U.S. population census is conducted every 10 years.

An **experiment** imposes a treatment on individuals in order to collect data on their response to the treatment. The treatment may be a medical treatment, or it can be any action that might affect a variable in the experiment, such as adding methanol to gasoline and then measuring its effect on fuel efficiency.

An **observational study** observes individuals and measures variables without controlling the individuals or their environment. This type of study is used when it is difficult to control or isolate the variable being studied, or when it may be unethical to subject people to a certain treatment or to withhold it from them.

A **survey** is an investigation of one or more characteristics of a population. In a survey, every member of a sample is asked one or more questions.

A **simulation** uses a model to reproduce the conditions of a situation or process so that the simulated outcomes closely match the real-world outcomes. Simulations allow you to study situations that are impractical or dangerous to create in real life.

EXAMPLE 4 Identifying Methods of Data Collection

Identify the method of data collection each situation describes.

a. A researcher records whether people at a gas station use hand sanitizer.

b. A landscaper fertilizes 20 lawns with a regular fertilizer mix and 20 lawns with a new organic fertilizer. The landscaper then compares the lawns after 10 weeks and determines which fertilizer is better.

SOLUTION

a. The researcher is gathering data without controlling the individuals or applying a treatment. So, this situation is an *observational study.*

b. A treatment (organic fertilizer) is being applied to some of the individuals (lawns) in the study. So, this situation is an *experiment.*

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

Identify the method of data collection the situation describes.

5. Members of a student council at your school ask every eighth student who enters the cafeteria whether they like the snacks in the school's vending machines.

6. A park ranger measures and records the heights of trees in a park as they grow.

7. A researcher uses a computer program to help determine how fast an influenza virus might spread within a city.

STUDY TIP

Bias may also be introduced in survey questioning in other ways, such as by the order in which questions are asked or by respondents giving answers they believe will please the questioner.

Recognizing Bias in Survey Questions

When designing a survey, it is important to word survey questions so they do not lead to biased results. Answers to poorly worded questions may not accurately reflect the opinions or actions of those being surveyed. Questions that are flawed in a way that leads to inaccurate results are called **biased questions**. Avoid questions that:

- encourage a particular response
- do not provide enough information to give an accurate opinion
- are too sensitive to answer truthfully
- address more than one issue

EXAMPLE 5 Identify and Correct Bias in Survey Questioning

A dentist surveys his patients by asking, "Do you brush your teeth at least twice per day and floss every day?" Explain why the question may be biased or otherwise introduce bias into the survey. Then describe a way to correct the flaw.

SOLUTION

Patients who brush less than twice per day or do not floss daily may be afraid to admit this because the dentist is asking the question. One improvement may be to have patients answer questions about dental hygiene on paper and then put the paper anonymously into a box.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

8. Explain why the survey question below may be biased or otherwise introduce bias into the survey. Then describe a way to correct the flaw.

"Do you agree that our school cafeteria should switch to a healthier menu?"

11.4 Lesson

Core Vocabulary

controlled experiment, *p. 620*
control group, *p. 620*
treatment group, *p. 620*
randomization, *p. 620*
randomized comparative experiment, *p. 620*
placebo, *p. 620*
replication, *p. 622*

Previous
sample size

What You Will Learn

- Describe experiments.
- Recognize how randomization applies to experiments and observational studies.
- Analyze experimental designs.

Describing Experiments

In a **controlled experiment**, two groups are studied under identical conditions with the exception of one variable. The group under ordinary conditions that is subjected to no treatment is the **control group**. The group that is subjected to the treatment is the **treatment group**.

Randomization is a process of randomly assigning subjects to different treatment groups. In a **randomized comparative experiment**, subjects are randomly assigned to the control group or the treatment group. In some cases, subjects in the control group are given a **placebo**, which is a harmless, unmedicated treatment that resembles the actual treatment. The comparison of the control group and the treatment group makes it possible to determine any effects of the treatment.

Randomization minimizes bias and produces groups of individuals who are theoretically similar in all ways before the treatment is applied. Conclusions drawn from an experiment that is not a randomized comparative experiment may not be valid.

EXAMPLE 1 Evaluating Published Reports

Determine whether each study is a randomized comparative experiment. If it is, describe the treatment, the treatment group, and the control group. If it is not, explain why not and discuss whether the conclusions drawn from the study are valid.

a. *Health Watch*

Vitamin C Lowers Cholesterol

At a health clinic, patients were given the choice of whether to take a dietary supplement of 500 milligrams of vitamin C each day. Fifty patients who took the supplement were monitored for one year, as were 50 patients who did not take the supplement. At the end of one year, patients who took the supplement had 15% lower cholesterol levels than patients in the other group.

b. *Supermarket Checkout*

Check Out Even Faster

To test the new design of its self checkout, a grocer gathered 142 customers and randomly divided them into two groups. One group used the new self checkout and one group used the old self checkout to buy the same groceries. Users of the new self checkout were able to complete their purchases 16% faster.

STUDY TIP

The study in part (a) is an *observational study* because the treatment is not being imposed.

SOLUTION

a. The study is not a randomized comparative experiment because the individuals were not randomly assigned to a control group and a treatment group. The conclusion that vitamin C lowers cholesterol may or may not be valid. There may be other reasons why patients who took the supplement had lower cholesterol levels. For instance, patients who voluntarily take the supplement may be more likely to have other healthy eating or lifestyle habits that could affect their cholesterol levels.

b. The study is a randomized comparative experiment. The treatment is the use of the new self checkout. The treatment group is the individuals who use the new self checkout. The control group is the individuals who use the old self checkout.

Monitoring Progress Help in English and Spanish at *BigIdeasMath.com*

Motorist News

Early Birds Make Better Drivers

A recent study shows that adults who rise before 6:30 A.M. are better drivers than other adults. The study monitored the driving records of 140 volunteers who always wake up before 6:30 and 140 volunteers who never wake up before 6:30. The early risers had 12% fewer accidents.

1. Determine whether the study is a randomized comparative experiment. If it is, describe the treatment, the treatment group, and the control group. If it is not, explain why not and discuss whether the conclusions drawn from the study are valid.

Randomization in Experiments and Observational Studies

You have already learned about random sampling and its usefulness in surveys. Randomization applies to experiments and observational studies as shown below.

Experiment	Observational study
Individuals are assigned at random to the treatment group or the control group.	When possible, random samples can be selected for the groups being studied.

Good experiments and observational studies are designed to compare data from two or more groups and to show any relationship between variables. Only a well-designed *experiment*, however, can determine a cause-and-effect relationship.

Core Concept

Comparative Studies and Causality

- A rigorous randomized comparative experiment, by eliminating sources of variation other than the controlled variable, can make valid cause-and-effect conclusions possible.
- An observational study can identify *correlation* between variables, but not *causality*. Variables, other than what is being measured, may be affecting the results.

EXAMPLE 2 **Designing an Experiment or Observational Study**

Explain whether the following research topic is best investigated through an experiment or an observational study. Then describe the design of the experiment or observational study.

You want to know whether vigorous exercise in older people results in longer life.

SOLUTION

The treatment, vigorous exercise, is not possible for those people who are already unhealthy, so it is not ethical to assign individuals to a control or treatment group. Use an observational study. Randomly choose one group of individuals who already exercise vigorously. Then randomly choose one group of individuals who do not exercise vigorously. Monitor the ages of the individuals in both groups at regular intervals. Note that because you are using an observational study, you should be able to identify a *correlation* between vigorous exercise in older people and longevity, but not *causality*.

Monitoring Progress Help in English and Spanish at *BigIdeasMath.com*

2. Determine whether the following research topic is best investigated through an experiment or an observational study. Then describe the design of the experiment or observational study.

   You want to know whether flowers sprayed twice per day with a mist of water stay fresh longer than flowers that are not sprayed.

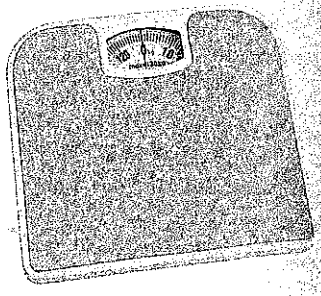
Analyzing Experimental Designs

An important part of experimental design is *sample size*, or the number of subjects in the experiment. To improve the validity of the experiment, **replication** is required, which is repetition of the experiment under the same or similar conditions.

UNDERSTANDING MATHEMATICAL TERMS

The *validity* of an experiment refers to the reliability of the results. The results of a valid experiment are more likely to be accepted.

EXAMPLE 3 Analyzing Experimental Designs



A pharmaceutical company wants to test the effectiveness of a new chewing gum designed to help people lose weight. Identify a potential problem, if any, with each experimental design. Then describe how you can improve it.

a. The company identifies 10 people who are overweight. Five subjects are given the new chewing gum and the other 5 are given a placebo. After 3 months, each subject is evaluated and it is determined that the 5 subjects who have been using the new chewing gum have lost weight.

b. The company identifies 10,000 people who are overweight. The subjects are divided into groups according to gender. Females receive the new chewing gum and males receive the placebo. After 3 months, a significantly large number of the female subjects have lost weight.

c. The company identifies 10,000 people who are overweight. The subjects are divided into groups according to age. Within each age group, subjects are randomly assigned to receive the new chewing gum or the placebo. After 3 months, a significantly large number of the subjects who received the new chewing gum have lost weight.

STUDY TIP

The experimental design described in part (c) is an example of *randomized block design*.

SOLUTION

a. The sample size is not large enough to produce valid results. To improve the validity of the experiment, the sample size must be larger and the experiment must be replicated.

b. Because the subjects are divided into groups according to gender, the groups are not similar. The new chewing gum may have more of an effect on women than on men, or more of an effect on men than on women. It is not possible to see such an effect with the experiment the way it is designed. The subjects can be divided into groups according to gender, but within each group, they must be randomly assigned to the treatment group or the control group.

c. The subjects are divided into groups according to a similar characteristic (age). Because subjects within each age group are randomly assigned to receive the new chewing gum or the placebo, replication is possible.

Monitoring Progress Help in English and Spanish at *BigIdeasMath.com*

3. In Example 3, the company identifies 250 people who are overweight. The subjects are randomly assigned to a treatment group or a control group. In addition, each subject is given a DVD that documents the dangers of obesity. After 3 months, most of the subjects placed in the treatment group have lost weight. Identify a potential problem with the experimental design. Then describe how you can improve it.

4. You design an experiment to test the effectiveness of a vaccine against a strain of influenza. In the experiment, 100,000 people receive the vaccine and another 100,000 people receive a placebo. Identify a potential problem with the experimental design. Then describe how you can improve it.