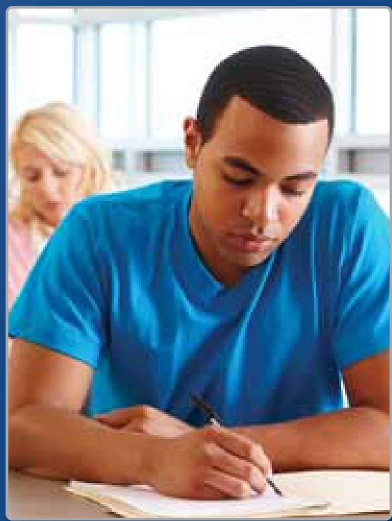
11 Data Analysis and Statistics

Solar Power ***(p. 631)***

Reading ***(p. 624)***

Volcano Damage ***(p. 615)***



SAT Scores ***(p. 605)***

Infant Weights ***(p. 598)***

Maintaining Mathematical Proficiency

Comparing Measures of Center

Example 1 **Find the mean, median, and mode of the data set 4, 11, 16, 8, 9, 40, 4, 12, 13, 5, and 10. Then determine which measure of center best represents the data. Explain.**

Mean $\bar{x} = \dfrac{4 + 11 + 16 + 8 + 9 + 40 + 4 + 12 + 13 + 5 + 10}{11} = 12$

Median 4, 4, 5, 8, 9, 10, 11, 12, 13, 16, 40 — Order the data. The middle value is 10.

Mode 4, 4, 5, 8, 9, 10, 11, 12, 13, 16, 40 — 4 occurs most often.

▶ The mean is 12, the median is 10, and the mode is 4. The median best represents the data. The mode is less than most of the data, and the mean is greater than most of the data.

Find the mean, median, and mode of the data set. Then determine which measure of center best represents the data. Explain.

1. 36, 82, 94, 83, 86, 82

2. 74, 89, 71, 70, 68, 70

3. 1, 18, 12, 16, 11, 15, 17, 44, 44

Finding a Standard Deviation

Example 2 **Find and interpret the standard deviation of the data set 10, 2, 6, 8, 12, 15, 18, and 25. Use a table to organize your work.**

x	$\bar{x}$	$x - \bar{x}$	$(x - \bar{x})^2$
10	12	−2	4
2	12	−10	100
6	12	−6	36
8	12	−4	16
12	12	0	0
15	12	3	9
18	12	6	36
25	12	13	169

Step 1 Find the mean, $\bar{x}$.

$$\bar{x} = \frac{96}{8} = 12$$

Step 2 Find the deviation of each data value, $x - \bar{x}$, as shown in the table.

Step 3 Square each deviation, $(x - \bar{x})^2$, as shown in the table.

Step 4 Find the mean of the squared deviations.

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{4 + 100 + \cdots + 169}{8} = \frac{370}{8} = 46.25$$

Step 5 Use a calculator to take the square root of the mean of the squared deviations.

$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}} = \sqrt{\frac{370}{8}} = \sqrt{46.25} \approx 6.80$$

▶ The standard deviation is about 6.80. This means that the typical data value differs from the mean by about 6.80 units.

Find and interpret the standard deviation of the data set.

4. 43, 48, 41, 51, 42

5. 28, 26, 21, 44, 29, 32

6. 65, 56, 49, 66, 62, 52, 53, 49

7. **ABSTRACT REASONING** Describe a data set that has a standard deviation of zero. Can a standard deviation be negative? Explain your reasoning.

Mathematical Practices

Mathematically proficient students use diagrams and graphs to show relationships between data. They also analyze data to draw conclusions.

Modeling with Mathematics

Core Concept

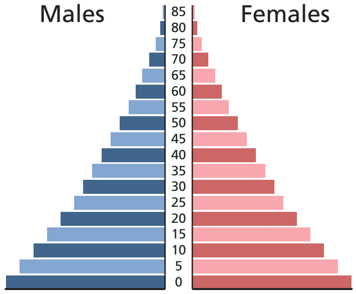
Information Design

Information design is the designing of data and information so it can be understood and used. Throughout this book, you have seen several types of information design. In the modern study of statistics, many types of designs require technology to analyze the data and organize the graphical design.

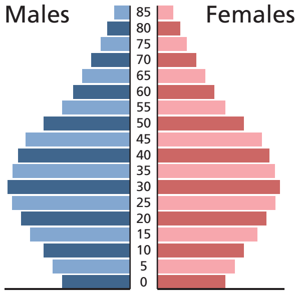
EXAMPLE 1 Comparing Age Pyramids

You can use an *age pyramid* to compare the ages of males and females in the population of a country. Compare the mean, median, and mode of each age pyramid.

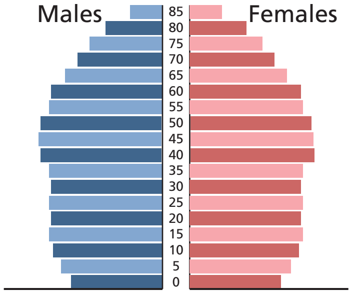
a.



b.



c.



SOLUTION

a. The relative frequency of each successive age group (from 0–4 to 85+) is less than the preceding age group. The mean is roughly 25 years, the median is roughly 20 years, and the mode is the youngest age group, 0–4 years.

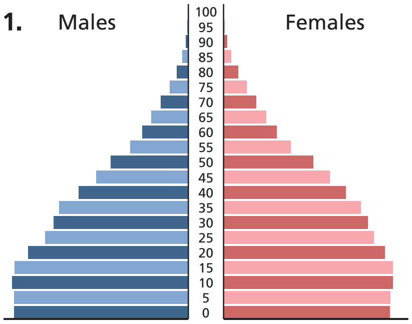
b. The mean, median, and mode are all roughly 32 years.

c. The mean, median, and mode are all roughly middle age, around 40 or 45 years.

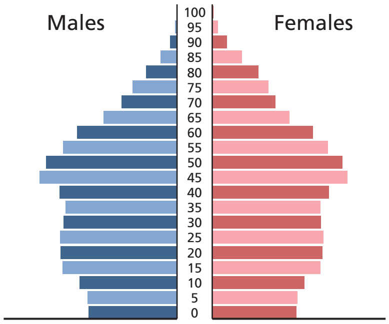
Monitoring Progress

Use the Internet or some other reference to determine which age pyramid is that of Canada, Japan, and Mexico. Compare the mean, median, and mode of the three age pyramids.

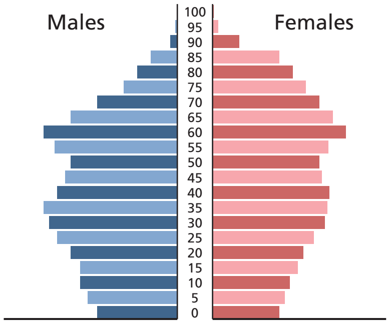
1.



2.



3.



11.1 Using Normal Distributions

Essential Question In a normal distribution, about what percent of the data lies within one, two, and three standard deviations of the mean?

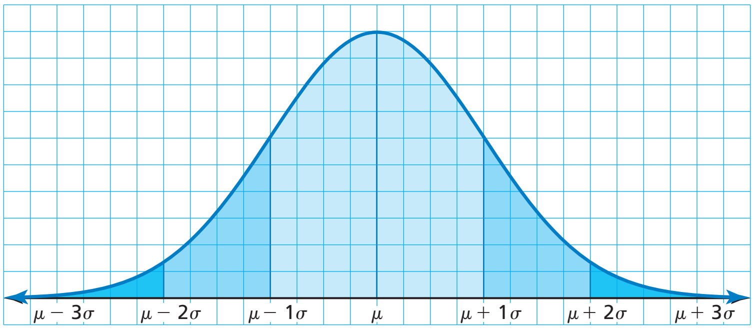
Recall that the standard deviation σ of a numerical data set is given by

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2}{n}}$$

where n is the number of values in the data set and μ is the mean of the data set.

EXPLORATION 1 Analyzing a Normal Distribution

Work with a partner. In many naturally occurring data sets, the histogram of the data is bell-shaped. In statistics, such data sets are said to have a *normal distribution*. For the normal distribution shown below, estimate the percent of the data that lies within one, two, and three standard deviations of the mean. Each square on the grid represents 1%.



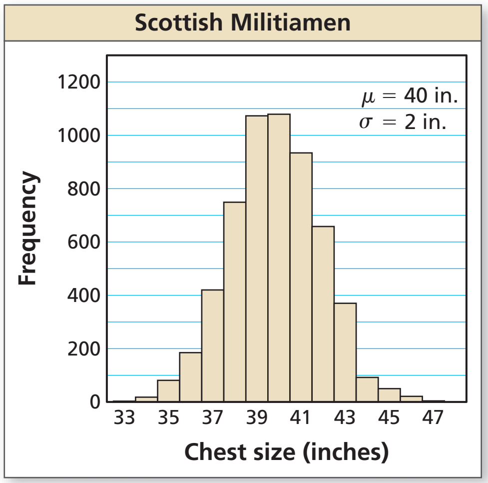
EXPLORATION 2 Analyzing a Data Set

Work with a partner. A famous data set was collected in Scotland in the mid-1800s. It contains the chest sizes (in inches) of 5738 men in the Scottish Militia. Do the data fit a normal distribution? Explain.

MODELING WITH MATHEMATICS

To be proficient in math, you need to analyze relationships mathematically to draw conclusions.

Chest size	Number of men
33	3
34	18
35	81
36	185
37	420
38	749
39	1073
40	1079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1



Communicate Your Answer

3. In a normal distribution, about what percent of the data lies within one, two, and three standard deviations of the mean?
4. Use the Internet or some other reference to find another data set that is normally distributed. Display your data in a histogram.

11.1 Lesson

Core Vocabulary

normal distribution, *p. 596*
normal curve, *p. 596*
standard normal distribution, *p. 597*
z-score, *p. 597*

Previous
probability distribution
symmetric
mean
standard deviation
skewed
median

What You Will Learn

- Calculate probabilities using normal distributions.
- Use *z*-scores and the standard normal table to find probabilities.
- Recognize data sets that are normal.

Normal Distributions

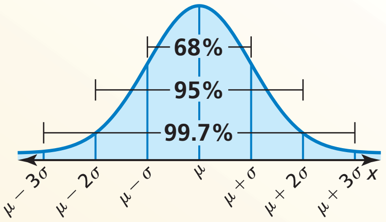
You have studied probability distributions. One type of probability distribution is a *normal distribution*. The graph of a **normal distribution** is a bell-shaped curve called a **normal curve** that is symmetric about the mean.

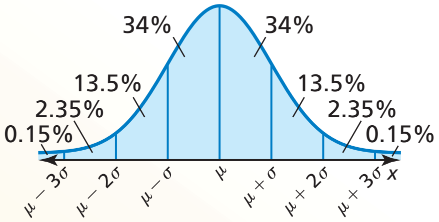
Core Concept

Areas Under a Normal Curve

A normal distribution with mean μ (the Greek letter *mu*) and standard deviation σ (the Greek letter *sigma*) has these properties.

- The total area under the related normal curve is 1.
- About 68% of the area lies within 1 standard deviation of the mean.
- About 95% of the area lies within 2 standard deviations of the mean.
- About 99.7% of the area lies within 3 standard deviations of the mean.



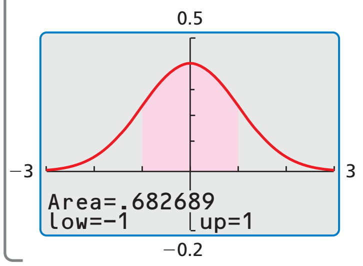


From the second bulleted statement above and the symmetry of a normal curve, you can deduce that 34% of the area lies within 1 standard deviation to the left of the mean, and 34% of the area lies within 1 standard deviation to the right of the mean. The second diagram above shows other partial areas based on the properties of a normal curve.

The areas under a normal curve can be interpreted as probabilities in a normal distribution. So, in a normal distribution, the probability that a randomly chosen x-value is between a and b is given by the area under the normal curve between a and b.

USING A GRAPHING CALCULATOR

A graphing calculator can be used to find areas under normal curves. For example, the normal distribution shown below has mean 0 and standard deviation 1. The graphing calculator screen shows that the area within 1 standard deviation of the mean is about 0.68, or 68%.

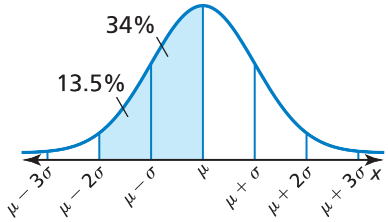


EXAMPLE 1 Finding a Normal Probability

A normal distribution has mean μ and standard deviation σ. An x-value is randomly selected from the distribution. Find $P(\mu - 2\sigma \leq x \leq \mu)$.

SOLUTION

The probability that a randomly selected x-value lies between $\mu - 2\sigma$ and μ is the shaded area under the normal curve shown.



$P(\mu - 2\sigma \leq x \leq \mu) = 0.135 + 0.34 = 0.475$

EXAMPLE 2 Interpreting Normally Distributed Data

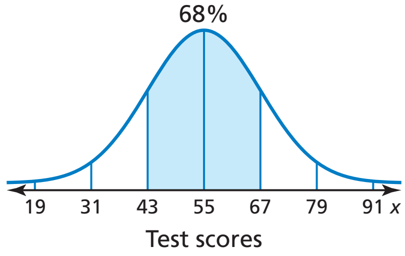
The scores for a state's peace officer standards and training test are normally distributed with a mean of 55 and a standard deviation of 12. The test scores range from 0 to 100.

a. About what percent of the people taking the test have scores between 43 and 67?

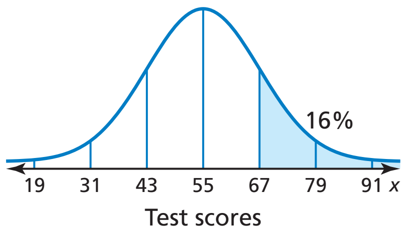
b. An agency in the state will only hire applicants with test scores of 67 or greater. About what percent of the people have test scores that make them eligible to be hired by the agency?

SOLUTION

a. The scores of 43 and 67 represent one standard deviation on either side of the mean, as shown. So, about 68% of the people taking the test have scores between 43 and 67.

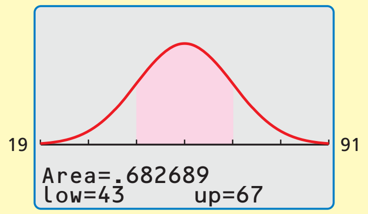


b. A score of 67 is one standard deviation to the right of the mean, as shown. So, the percent of the people who have test scores that make them eligible to be hired by the agency is about 13.5% + 2.35% + 0.15%, or 16%.

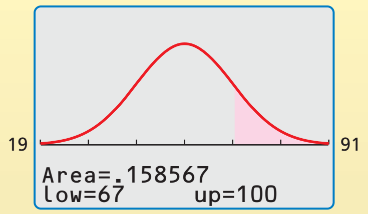


Check

a.



b.



Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

A normal distribution has mean μ and standard deviation σ. Find the indicated probability for a randomly selected x-value from the distribution.

1. $P(x \le \mu)$

2. $P(x \ge \mu)$

3. $P(\mu \le x \le \mu + 2\sigma)$

4. $P(\mu - \sigma \le x \le \mu)$

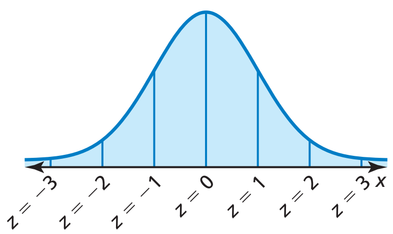
5. $P(x \le \mu - 3\sigma)$

6. $P(x \ge \mu + \sigma)$

7. **WHAT IF?** In Example 2, about what percent of the people taking the test have scores between 43 and 79?

The Standard Normal Distribution

The **standard normal distribution** is the normal distribution with mean 0 and standard deviation 1. The formula below can be used to transform x-values from a normal distribution with mean μ and standard deviation σ into z-values having a standard normal distribution.



Formula $z = \dfrac{x - \mu}{\sigma}$

Subtract the mean from the given x-value, then divide by the standard deviation.

The z-value for a particular x-value is called the ***z*-score** for the x-value and is the number of standard deviations the x-value lies above or below the mean μ.

READING

In the table, the value .0000+ means "slightly more than 0" and the value 1.0000− means "slightly less than 1."

For a randomly selected z-value from a standard normal distribution, you can use the table below to find the probability that z is less than or equal to a given value. For example, the table shows that $P(z \leq -0.4) = 0.3446$. You can find the value of $P(z \leq -0.4)$ in the table by finding the value where row −0 and column .4 intersect.

Standard Normal Table

z	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
−3	.0013	.0010	.0007	.0005	.0003	.0002	.0002	.0001	.0001	.0000+
−2	.0228	.0179	.0139	.0107	.0082	.0062	.0047	.0035	.0026	.0019
−1	.1587	.1357	.1151	.0968	.0808	.0668	.0548	.0446	.0359	.0287
−0	.5000	.4602	.4207	.3821	.3446	.3085	.2743	.2420	.2119	.1841
0	.5000	.5398	.5793	.6179	.6554	.6915	.7257	.7580	.7881	.8159
1	.8413	.8643	.8849	.9032	.9192	.9332	.9452	.9554	.9641	.9713
2	.9772	.9821	.9861	.9893	.9918	.9938	.9953	.9965	.9974	.9981
3	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000−

You can also use the standard normal table to find probabilities for any normal distribution by first converting values from the distribution to z-scores.

EXAMPLE 3 Using a *z*-Score and the Standard Normal Table

A study finds that the weights of infants at birth are normally distributed with a mean of 3270 grams and a standard deviation of 600 grams. An infant is randomly chosen. What is the probability that the infant weighs 4170 grams or less?

SOLUTION

Step 1 Find the z-score corresponding to an x-value of 4170.

$$z = \frac{x - \mu}{\sigma} = \frac{4170 - 3270}{600} = 1.5$$

Step 2 Use the table to find $P(z \leq 1.5)$. The table shows that $P(z \leq 1.5) = 0.9332$.

Standard Normal Table

z	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
−3	.0013	.0010	.0007	.0005	.0003	.0002	.0002	.0001	.0001	.0000+
−2	.0228	.0179	.0139	.0107	.0082	.0062	.0047	.0035	.0026	.0019
−1	.1587	.1357	.1151	.0968	.0808	.0668	.0548	.0446	.0359	.0287
−0	.5000	.4602	.4207	.3821	.3446	.3085	.2743	.2420	.2119	.1841
0	.5000	.5398	.5793	.6179	.6554	.6915	.7257	.7580	.7881	.8159
1	.8413	.8643	.8849	.9032	.9192	.9332	.9452	.9554	.9641	.9713

STUDY TIP

When n% of the data are less than or equal to a certain value, that value is called the nth *percentile*. In Example 3, a weight of 4170 grams is the 93rd percentile.

So, the probability that the infant weighs 4170 grams or less is about 0.9332.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

8. **WHAT IF?** In Example 3, what is the probability that the infant weighs 3990 grams or more?

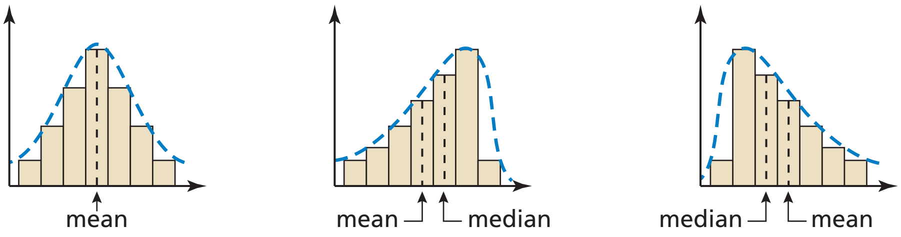
9. Explain why it makes sense that $P(z \leq 0) = 0.5$.

Recognizing Normal Distributions

Not all distributions are normal. For instance, consider the histograms shown below. The first histogram has a normal distribution. Notice that it is bell-shaped and symmetric. Recall that a distribution is symmetric when you can draw a vertical line that divides the histogram into two parts that are mirror images. Some distributions are skewed. The second histogram is *skewed left* and the third histogram is *skewed right*. The second and third histograms do *not* have normal distributions.

UNDERSTANDING MATHEMATICAL TERMS

Be sure you understand that you cannot use a normal distribution to interpret skewed distributions. The areas under a normal curve do not correspond to the areas of a skewed distribution.



Bell-shaped and symmetric
- histogram has a normal distribution
- mean = median

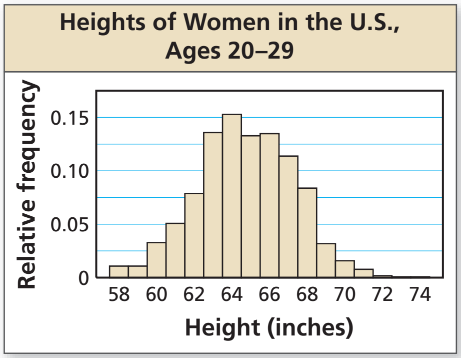
Skewed left
- histogram does not have a normal distribution
- mean < median

Skewed right
- histogram does not have a normal distribution
- mean > median

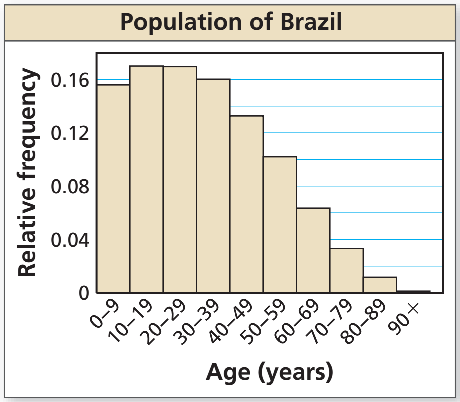
EXAMPLE 4 Recognizing Normal Distributions

Determine whether each histogram has a normal distribution.

a.



b.



SOLUTION

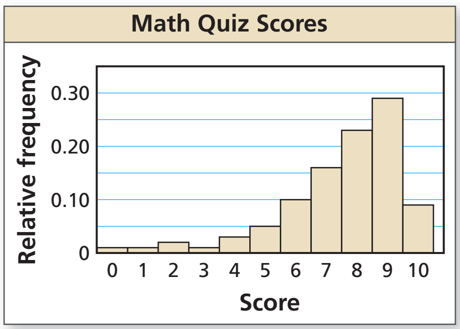
a. The histogram is bell-shaped and fairly symmetric. So, the histogram has an approximately normal distribution.

b. The histogram is skewed right. So, the histogram does not have a normal distribution, and you cannot use a normal distribution to interpret the histogram.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

10. Determine whether the histogram has a normal distribution.



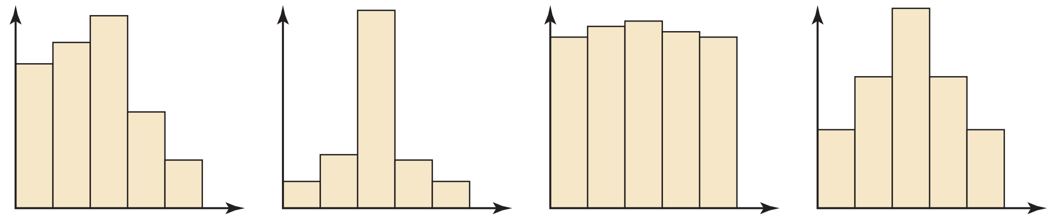
11.1 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **WRITING** Describe how to use the standard normal table to find $P(z \leq 1.4)$.

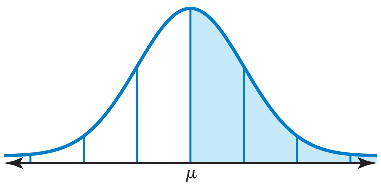
2. **WHICH ONE DOESN'T BELONG?** Which histogram does *not* belong with the other three? Explain your reasoning.



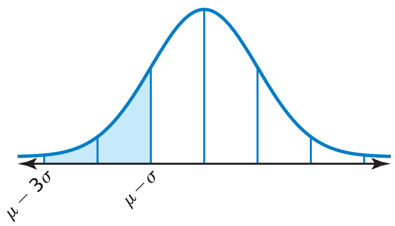
Monitoring Progress and Modeling with Mathematics

ATTENDING TO PRECISION **In Exercises 3–6, give the percent of the area under the normal curve represented by the shaded region(s).**

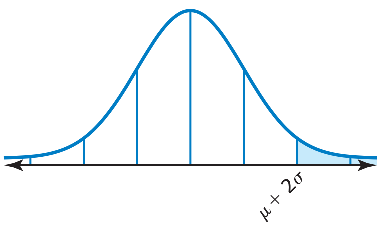
3.



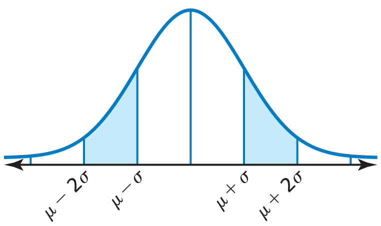
4.



5.



6.



In Exercises 7–12, a normal distribution has mean μ and standard deviation σ. Find the indicated probability for a randomly selected x-value from the distribution. *(See Example 1.)*

7. $P(x \leq \mu - \sigma)$

8. $P(x \geq \mu - \sigma)$

9. $P(x \geq \mu + 2\sigma)$

10. $P(x \leq \mu + \sigma)$

11. $P(\mu - \sigma \leq x \leq \mu + \sigma)$

12. $P(\mu - 3\sigma \leq x \leq \mu)$

In Exercises 13–18, a normal distribution has a mean of 33 and a standard deviation of 4. Find the probability that a randomly selected x-value from the distribution is in the given interval.

13. between 29 and 37

14. between 33 and 45

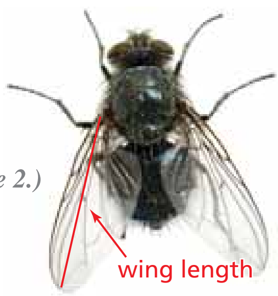
15. at least 25

16. at least 29

17. at most 37

18. at most 21

19. **PROBLEM SOLVING** The wing lengths of houseflies are normally distributed with a mean of 4.6 millimeters and a standard deviation of 0.4 millimeter. *(See Example 2.)*



   a. About what percent of houseflies have wing lengths between 3.8 millimeters and 5.0 millimeters?

   b. About what percent of houseflies have wing lengths longer than 5.8 millimeters?

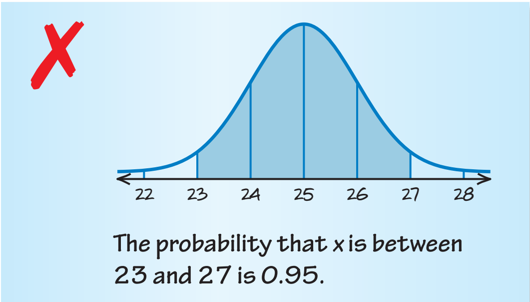
20. **PROBLEM SOLVING** The times a fire department takes to arrive at the scene of an emergency are normally distributed with a mean of 6 minutes and a standard deviation of 1 minute.

a. For about what percent of emergencies does the fire department arrive at the scene in 8 minutes or less?

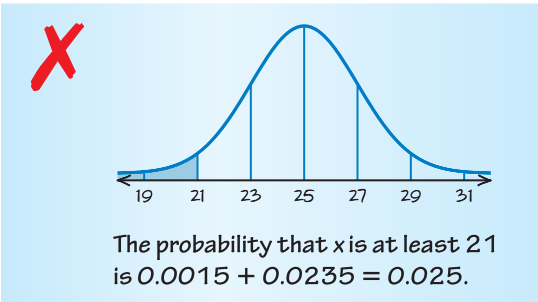
b. The goal of the fire department is to reach the scene of an emergency in 5 minutes or less. About what percent of the time does the fire department achieve its goal?

ERROR ANALYSIS **In Exercises 21 and 22, a normal distribution has a mean of 25 and a standard deviation of 2. Describe and correct the error in finding the probability that a randomly selected x-value is in the given interval.**

21. between 23 and 27



22. at least 21

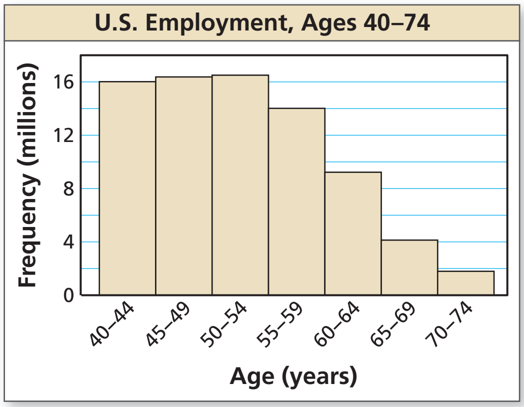


23. **PROBLEM SOLVING** A busy time to visit a bank is during its Friday evening rush hours. For these hours, the waiting times at the drive-through window are normally distributed with a mean of 8 minutes and a standard deviation of 2 minutes. You have no more than 11 minutes to do your banking and still make it to your meeting on time. What is the probability that you will be late for the meeting? *(See Example 3.)*

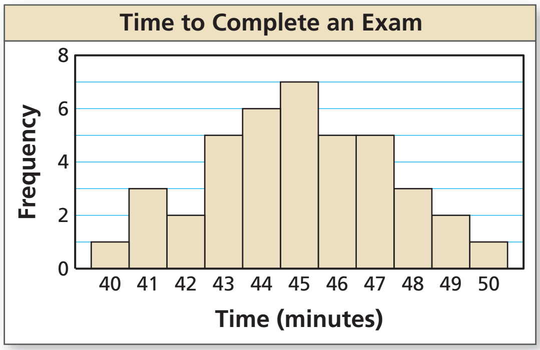
24. **PROBLEM SOLVING** Scientists conducted aerial surveys of a seal sanctuary and recorded the number x of seals they observed during each survey. The numbers of seals observed were normally distributed with a mean of 73 seals and a standard deviation of 14.1 seals. Find the probability that at most 50 seals were observed during a randomly chosen survey.

In Exercises 25 and 26, determine whether the histogram has a normal distribution. *(See Example 4.)*

25.



26.



27. **ANALYZING RELATIONSHIPS** The table shows the numbers of tickets that are sold for various baseball games in a league over an entire season. Display the data in a histogram. Do the data fit a normal distribution? Explain.

Tickets sold	Frequency
150–189	1
190–229	2
230–269	4
270–309	8
310–349	8
350–389	7

28. **PROBLEM SOLVING** The guayule plant, which grows in the southwestern United States and in Mexico, is one of several plants that can be used as a source of rubber. In a large group of guayule plants, the heights of the plants are normally distributed with a mean of 12 inches and a standard deviation of 2 inches.

a. What percent of the plants are taller than 16 inches?

b. What percent of the plants are at most 13 inches?

c. What percent of the plants are between 7 inches and 14 inches?

d. What percent of the plants are at least 3 inches taller than or at least 3 inches shorter than the mean height?

29. **REASONING** Boxes of cereal are filled by a machine. Tests show that the amount of cereal in each box varies. The weights are normally distributed with a mean of 20 ounces and a standard deviation of 0.25 ounce. Four boxes of cereal are randomly chosen.

a. What is the probability that all four boxes contain no more than 19.4 ounces of cereal?

b. Do you think the machine is functioning properly? Explain.

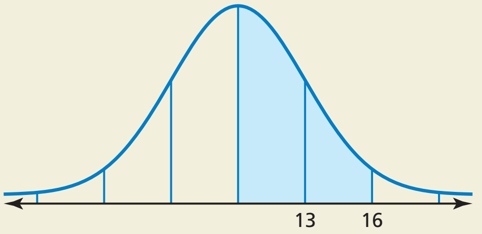
30. **THOUGHT PROVOKING** Sketch the graph of the standard normal distribution function, given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Estimate the area of the region bounded by the x-axis, the graph of f, and the vertical lines $x = -3$ and $x = 3$.

31. **REASONING** For normally distributed data, describe the value that represents the 84th percentile in terms of the mean and standard deviation.

32. **HOW DO YOU SEE IT?** In the figure, the shaded region represents 47.5% of the area under a normal curve. What are the mean and standard deviation of the normal distribution?



33. **DRAWING CONCLUSIONS** You take both the SAT (Scholastic Aptitude Test) and the ACT (American College Test). You score 650 on the mathematics section of the SAT and 29 on the mathematics section of the ACT. The SAT test scores and the ACT test scores are each normally distributed. For the SAT, the mean is 514 and the standard deviation is 118. For the ACT, the mean is 21.0 and the standard deviation is 5.3.

a. What percentile is your SAT math score?

b. What percentile is your ACT math score?

c. On which test did you perform better? Explain your reasoning.

34. **WRITING** Explain how you can convert ACT scores into corresponding SAT scores when you know the mean and standard deviation of each distribution.

35. **MAKING AN ARGUMENT** A data set has a median of 80 and a mean of 90. Your friend claims that the distribution of the data is skewed left. Is your friend correct? Explain your reasoning.

36. **CRITICAL THINKING** The average scores on a statistics test are normally distributed with a mean of 75 and a standard deviation of 10. You randomly select a test score x. Find $P(|x - \mu| \geq 15)$.

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Graph the function. Identify the x-intercepts and the points where the local maximums and local minimums occur. Determine the intervals for which the function is increasing or decreasing. *(Section 4.8)*

37. $f(x) = x^3 - 4x^2 + 5$

38. $g(x) = \frac{1}{4}x^4 - 2x^2 - x - 3$

39. $h(x) = -0.5x^2 + 3x + 7$

40. $f(x) = -x^4 + 6x^2 - 13$

11.2 Populations, Samples, and Hypotheses

Essential Question How can you test theoretical probability using sample data?

EXPLORATION 1 Using Sample Data

Work with a partner.

a. When two six-sided dice are rolled, what is the theoretical probability that you roll the same number on both dice?

b. Conduct an experiment to check your answer in part (a). What sample size did you use? Explain your reasoning.

c. Use the dice rolling simulator at *BigIdeasMath.com* to complete the table and check your answer to part (a). What happens as you increase the sample size?

Number of Rolls	Number of Times Same Number Appears	Experimental Probability
100		
500		
1000		
5000		
10,000		

USING TOOLS STRATEGICALLY

To be proficient in math, you need to use technology to visualize the results of varying assumptions, explore consequences, and compare predictions with data.

EXPLORATION 2 Using Sample Data

Work with a partner.

a. When three six-sided dice are rolled, what is the theoretical probability that you roll the same number on all three dice?

b. Compare the theoretical probability you found in part (a) with the theoretical probability you found in Exploration 1(a).

c. Conduct an experiment to check your answer in part (a). How does adding a die affect the sample size that you use? Explain your reasoning.

d. Use the dice rolling simulator at *BigIdeasMath.com* to check your answer to part (a). What happens as you increase the sample size?

Communicate Your Answer

3. How can you test theoretical probability using sample data?

4. Conduct an experiment to determine the probability of rolling a sum of 7 when two six-sided dice are rolled. Then find the theoretical probability and compare your answers.

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

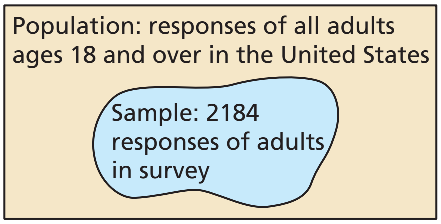
Identify the population and the sample. Describe the sample.

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

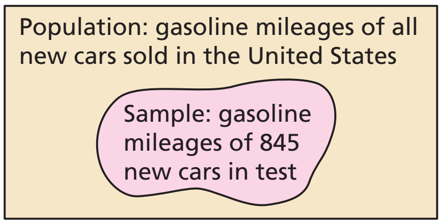
b. To estimate the gasoline mileage of new cars sold in the United States, a consumer advocacy group tests 845 new cars and finds they have an average of 25.1 miles per gallon.

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.

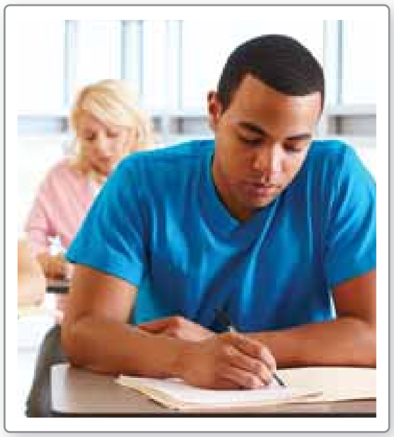


b. The population consists of the gasoline mileages of all new cars sold in the United States, and the sample consists of the gasoline mileages of the 845 new cars tested by the group. Notice in the diagram that the sample is a subset of the gasoline mileages of all new cars in the United States. The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

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%.

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.

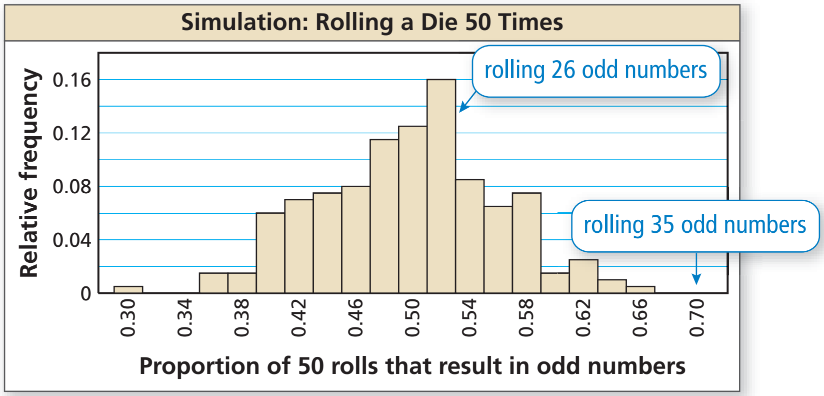
To analyze a hypothesis, you need to distinguish between results that can easily occur by chance and results that are highly unlikely to occur by chance. One way to analyze a hypothesis is to perform a *simulation*. When the results are highly unlikely to occur, the hypothesis is probably false.

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 or even 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 or even 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.

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.

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?

In Example 3(b), you concluded the maker's claim is probably false. In general, such conclusions may or may not be correct. 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

11.2 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **COMPLETE THE SENTENCE** A portion of a population that can be studied in order to make predictions about the entire population is a(n) ____________.

2. **WRITING** Describe the difference between a parameter and a statistic. Give an example of each.

3. **VOCABULARY** What is a hypothesis in statistics?

4. **WRITING** Describe two ways you can make an incorrect decision when analyzing a hypothesis.

Monitoring Progress and Modeling with Mathematics

In Exercises 5–8, determine whether the data are collected from a population or a sample. Explain your reasoning.

5. the number of high school students in the United States

6. the color of every third car that passes your house

7. a survey of 100 spectators at a sporting event with 1800 spectators

8. the age of each dentist in the United States

In Exercises 9–12, identify the population and sample. Describe the sample. *(See Example 1.)*

9. In the United States, a survey of 1152 adults ages 18 and over found that 403 of them pretend to use their smartphones to avoid talking to someone.

10. In the United States, a survey of 1777 adults ages 18 and over found that 1279 of them do some kind of spring cleaning every year.

11. In a school district, a survey of 1300 high school students found that 1001 of them like the new, healthy cafeteria food choices.

12. In the United States, a survey of 2000 households with at least one child found that 1280 of them eat dinner together every night.

In Exercises 13–16, determine whether the numerical value is a parameter or a statistic. Explain your reasoning. *(See Example 2.)*

13. The average annual salary of some physical therapists in a state is \$76,210.

14. In a recent year, 53% of the senators in the United States Senate were Democrats.

15. Seventy-three percent of all the students in a school would prefer to have school dances on Saturday.

16. A survey of U.S. adults found that 10% believe a cleaning product they use is not safe for the environment.

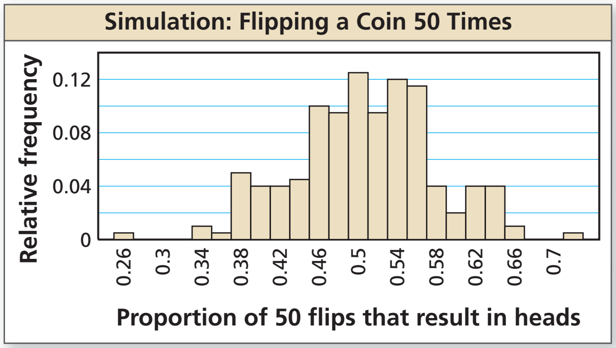
17. **ERROR ANALYSIS** A survey of 1270 high school students found that 965 students felt added stress because of their workload. Describe and correct the error in identifying the population and the sample.

✗ The population consists of all the students in the high school. The sample consists of the 965 students who felt added stress.

18. **ERROR ANALYSIS** Of all the players on a National Football League team, the mean age is 26 years. Describe and correct the error in determining whether the mean age represents a parameter or statistic.

✗ Because the mean age of 26 is based only on one football team, it is a statistic.

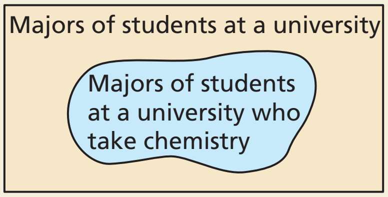
19. **MODELING WITH MATHEMATICS** You flip a coin 4 times and do not get a tails. You suspect this coin favors heads. The coin maker claims that the coin does not favor heads or tails. You simulate flipping the coin 50 times by repeatedly drawing 200 random samples of size 50. The histogram shows the results. What should you conclude when you flip the actual coin 50 times and get (a) 27 heads and (b) 33 heads? *(See Example 3.)*



20. **MODELING WITH MATHEMATICS** Use the histogram in Exercise 19 to determine what you should conclude when you flip the actual coin 50 times and get (a) 17 heads and (b) 23 heads.

21. **MAKING AN ARGUMENT** A random sample of five people at a movie theater from a population of 200 people gave the film 4 out of 4 stars. Your friend concludes that everyone in the movie theater would give the film 4 stars. Is your friend correct? Explain your reasoning.

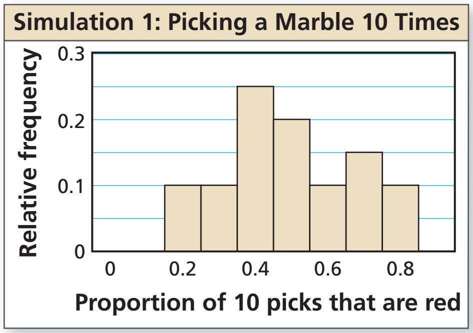
22. **HOW DO YOU SEE IT?** Use the Venn diagram to identify the population and sample. Explain your reasoning.

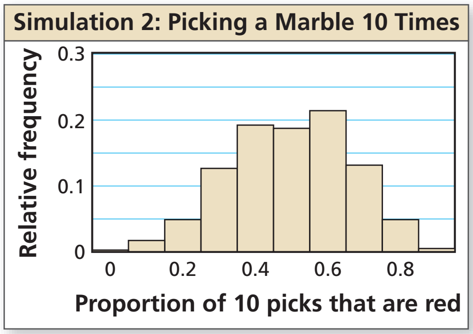


23. **OPEN-ENDED** Find a newspaper or magazine article that describes a survey. Identify the population and sample. Describe the sample.

24. **THOUGHT PROVOKING** You choose a random sample of 200 from a population of 2000. Each person in the sample is asked how many hours of sleep he or she gets each night. The mean of your sample is 8 hours. Is it possible that the mean of the entire population is only 7.5 hours of sleep each night? Explain.

25. **DRAWING CONCLUSIONS** You perform two simulations of repeatedly selecting a marble out of a bag with replacement that contains three red marbles and three blue marbles. The first simulation uses 20 random samples of size 10, and the second uses 400 random samples of size 10. The histograms show the results. Which simulation should you use to accurately analyze a hypothesis? Explain.





26. **PROBLEM SOLVING** You roll an eight-sided die five times and get a four every time. You suspect that the die favors the number four. The die maker claims that the die does not favor any number.

a. Perform a simulation involving 50 trials of rolling the actual die and getting a four to test the die maker's claim. Display the results in a histogram.

b. What should you conclude when you roll the actual die 50 times and get 20 fours? 7 fours?

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Solve the equation by completing the square. *(Section 3.3)*

27. $x^2 - 10x - 4 = 0$ **28.** $3t^2 + 6t = 18$ **29.** $s^2 + 10s + 8 = 0$

Solve the equation using the Quadratic Formula. *(Section 3.4)*

30. $n^2 + 2n + 2 = 0$ **31.** $4z^2 + 28z = 15$ **32.** $5w - w^2 = -11$

11.3 Collecting Data

Essential Question What are some considerations when undertaking a statistical study?

The goal of any statistical study is to collect data and then use the data to make a decision. Any decision you make using the results of a statistical study is only as reliable as the process used to obtain the data. If the process is flawed, then the resulting decision is questionable.

EXPLORATION 1 Analyzing Sampling Techniques

Work with a partner. Determine whether each sample is representative of the population. Explain your reasoning.

a. To determine the number of hours people exercise during a week, researchers use random-digit dialing and call 1500 people.

b. To determine how many text messages high school students send in a week, researchers post a survey on a website and receive 750 responses.

c. To determine how much money college students spend on clothes each semester, a researcher surveys 450 college students as they leave the university library.

d. To determine the quality of service customers receive, an airline sends an e-mail survey to each customer after the completion of a flight.

JUSTIFYING CONCLUSIONS

To be proficient in math, you need to justify your conclusions and communicate them to others.

EXPLORATION 2 Analyzing Survey Questions

Work with a partner. Determine whether each survey question is biased. Explain your reasoning. If so, suggest an unbiased rewording of the question.

a. Does eating nutritious, whole-grain foods improve your health?

b. Do you ever attempt the dangerous activity of texting while driving?

c. How many hours do you sleep each night?

d. How can the mayor of your city improve his or her public image?

EXPLORATION 3 Analyzing Survey Randomness and Truthfulness

Work with a partner. Discuss each potential problem in obtaining a random survey of a population. Include suggestions for overcoming the problem.

a. The people selected might not be a random sample of the population.

b. The people selected might not be willing to participate in the survey.

c. The people selected might not be truthful when answering the question.

d. The people selected might not understand the survey question.

Communicate Your Answer

4. What are some considerations when undertaking a statistical study?

5. Find a real-life example of a biased survey question. Then suggest an unbiased rewording of the question.

11.3 Lesson

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

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.

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.

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.

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.

STUDY TIP

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

STUDY TIP

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

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.

```
randInt(1,246)
                 84
                245
                 50
                197
                235
                 55
```

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.

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.

Core Concept

Methods of Collecting Data

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?"

11.3 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **VOCABULARY** Describe the difference between a stratified sample and a cluster sample.

2. **COMPLETE THE SENTENCE** A sample for which each member of a population has an equal chance of being selected is a(n) __________ sample.

3. **WRITING** Describe a situation in which you would use a simulation to collect data.

4. **WRITING** Describe the difference between an unbiased sample and a biased sample. Give one example of each.

Monitoring Progress and Modeling with Mathematics

In Exercises 5–8, identify the type of sample described. *(See Example 1.)*

5. The owners of a chain of 260 retail stores want to assess employee job satisfaction. Employees from 12 stores near the headquarters are surveyed.

6. Each employee in a company writes their name on a card and places it in a hat. The employees whose names are on the first two cards drawn each win a gift card.

7. A taxicab company wants to know whether its customers are satisfied with the service. Drivers survey every tenth customer during the day.

8. The owner of a community pool wants to ask patrons whether they think the water should be colder. Patrons are divided into four age groups, and a sample is randomly surveyed from each age group.

In Exercises 9–12, identify the type of sample and explain why the sample is biased. *(See Example 2.)*

9. A town council wants to know whether residents support having an off-leash area for dogs in the town park. Eighty dog owners are surveyed at the park.

10. A sportswriter wants to determine whether baseball coaches think wooden bats should be mandatory in collegiate baseball. The sportswriter mails surveys to all collegiate coaches and uses the surveys that are returned.

11. You want to find out whether booth holders at a convention were pleased with their booth locations. You divide the convention center into six sections and survey every booth holder in the fifth section.

12. Every tenth employee who arrives at a company health fair answers a survey that asks for opinions about new health-related programs.

13. **ERROR ANALYSIS** Surveys are mailed to every other household in a neighborhood. Each survey that is returned is used. Describe and correct the error in identifying the type of sample that is used.

> ✗ Because the surveys were mailed to every other household, the sample is a systematic sample.

14. **ERROR ANALYSIS** A researcher wants to know whether the U.S. workforce supports raising the minimum wage. Fifty high school students chosen at random are surveyed. Describe and correct the error in determining whether the sample is biased.

> ✗ Because the students were chosen at random, the sample is not biased.

In Exercises 15–18, determine whether the sample is biased. Explain your reasoning.

15. Every third person who enters an athletic event is asked whether he or she supports the use of instant replay in officiating the event.

16. A governor wants to know whether voters in the state support building a highway that will pass through a state forest. Business owners in a town near the proposed highway are randomly surveyed.

17. To assess customers' experiences making purchases online, a rating company e-mails purchasers and asks that they click on a link and complete a survey.

18. Your school principal randomly selects five students from each grade to complete a survey about classroom participation.

19. **WRITING** The staff of a student newsletter wants to conduct a survey of the students' favorite television shows. There are 1225 students in the school. Describe a method for selecting a random sample of 250 students to survey. *(See Example 3.)*

20. **WRITING** A national collegiate athletic association wants to survey 15 of the 120 head football coaches in a division about a proposed rules change. Describe a method for selecting a random sample of coaches to survey.

In Exercises 21–24, identify the method of data collection the situation describes. *(See Example 4.)*

21. A researcher uses technology to estimate the damage that will be done if a volcano erupts.

22. The owner of a restaurant asks 20 customers whether they are satisfied with the quality of their meals.

23. A researcher compares incomes of people who live in rural areas with those who live in large urban areas.

24. A researcher places bacteria samples in two different climates. The researcher then measures the bacteria growth in each sample after 3 days.

In Exercises 25–28, explain why the survey question may be biased or otherwise introduce bias into the survey. Then describe a way to correct the flaw. *(See Example 5.)*

25. "Do you agree that the budget of our city should be cut?"

26. "Would you rather watch the latest award-winning movie or just read some book?"

27. "The tap water coming from our western water supply contains twice the level of arsenic of water from our eastern supply. Do you think the government should address this health problem?"

28. A child asks, "Do you support the construction of a new children's hospital?"

In Exercises 29–32, determine whether the survey question may be biased or otherwise introduce bias into the survey. Explain your reasoning.

29. "Do you favor government funding to help prevent acid rain?"

30. "Do you think that renovating the old town hall would be a mistake?"

31. A police officer asks mall visitors, "Do you wear your seat belt regularly?"

32. "Do you agree with the amendments to the Clean Air Act?"

33. **REASONING** A researcher studies the effect of fiber supplements on heart disease. The researcher identified 175 people who take fiber supplements and 175 people who do not take fiber supplements. The study found that those who took the supplements had 19.6% fewer heart attacks. The researcher concludes that taking fiber supplements reduces the chance of heart attacks.

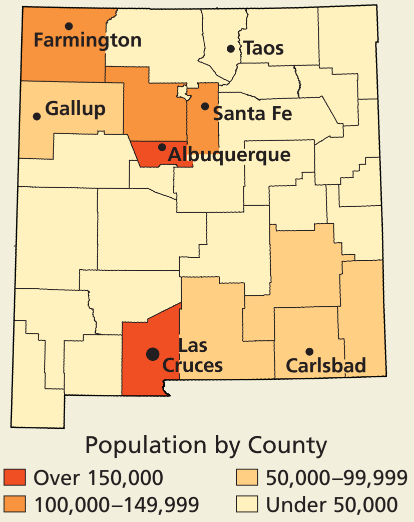
a. Explain why the researcher's conclusion may not be valid.

b. Describe how the researcher could have conducted the study differently to produce valid results.

34. HOW DO YOU SEE IT? A poll is conducted to predict the results of a statewide election in New Mexico before all the votes are counted. Fifty voters in each of the state's 33 counties are asked how they voted as they leave the polls.

a. Identify the type of sample described.

b. Explain how the diagram shows that the polling method could result in a biased sample.



35. WRITING Consider each type of sample listed on page 610. Which of the samples are most likely to lead to biased results? Explain.

36. THOUGHT PROVOKING What is the difference between a "blind experiment" and a "double-blind experiment?" Describe a possible advantage of the second type of experiment over the first.

37. WRITING A college wants to survey its graduating seniors to find out how many have already found jobs in their field of study after graduation.

a. What is the objective of the survey?

b. Describe the population for the survey.

c. Write two unbiased questions for the survey.

38. REASONING About 3.2% of U.S. adults follow a vegetarian-based diet. Two randomly selected groups of people were asked whether they follow such a diet. The first sample consists of 20 people and the second sample consists of 200 people. Which sample proportion is more likely to be representative of the national percentage? Explain.

39. MAKING AN ARGUMENT The U.S. Census is taken every 10 years to gather data from the population. Your friend claims that the sample cannot be biased. Is your friend correct? Explain.

40. OPEN-ENDED An airline wants to know whether travelers have enough leg room on its planes.

a. What method of data collection is appropriate for this situation?

b. Describe a sampling method that is likely to give biased results. Explain.

c. Describe a sampling method that is *not* likely to give biased results. Explain.

d. Write one biased question and one unbiased question for this situation.

41. REASONING A website contains a link to a survey that asks how much time each person spends on the Internet each week.

a. What type of sampling method is used in this situation?

b. Which population is likely to respond to the survey? What can you conclude?

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Evaluate the expression without using a calculator. *(Section 5.1)*

42. $4^{5/2}$ **43.** $27^{2/3}$ **44.** $-64^{1/3}$ **45.** $8^{-2/3}$

Simplify the expression. *(Section 5.2)*

46. $(4^{3/2} \cdot 4^{1/4})^4$ **47.** $(6^{1/3} \cdot 3^{1/3})^{-2}$ **48.** $\sqrt[3]{4} \cdot \sqrt[3]{16}$ **49.** $\dfrac{\sqrt[4]{405}}{\sqrt[4]{5}}$

11.1–11.3 What Did You Learn?

Core Vocabulary

normal distribution, *p. 596*
normal curve, *p. 596*
standard normal distribution, *p. 597*
z-score, *p. 597*
population, *p. 604*
sample, *p. 604*
parameter, *p. 605*
statistic, *p. 605*
hypothesis, *p. 605*
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*

Core Concepts

Section 11.1

Areas Under a Normal Curve, *p. 596*
Using *z*-Scores and the Standard Normal Table, *p. 597*
Recognizing Normal Distributions, *p. 599*

Section 11.2

Distinguishing Between Populations and Samples, *p. 604*
Analyzing Hypotheses, *p. 606*

Section 11.3

Types of Samples, *p. 610*
Methods of Collecting Data, *p. 612*

Mathematical Practices

1. What previously established results, if any, did you use to solve Exercise 31 on page 602?
2. What external resources, if any, did you use to answer Exercise 36 on page 616?

Study Skills

Reworking Your Notes

It's almost impossible to write down in your notes all the detailed information you are taught in class. A good way to reinforce the concepts and put them into your long-term memory is to rework your notes. When you take notes, leave extra space on the pages. You can go back after class and fill in:

- important definitions and rules
- additional examples
- questions you have about the material

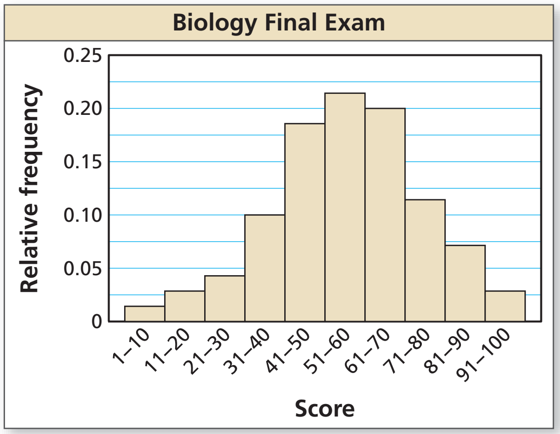
11.1–11.3 Quiz

A normal distribution has a mean of 32 and a standard deviation of 4. Find the probability that a randomly selected *x*-value from the distribution is in the given interval. *(Section 11.1)*

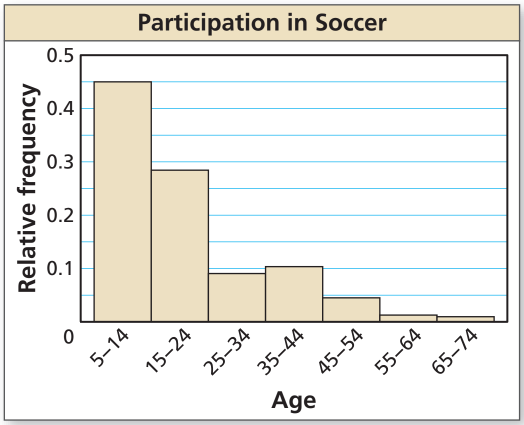
1. at least 28 **2.** between 20 and 32 **3.** at most 26 **4.** at most 35

Determine whether the histogram has a normal distribution. *(Section 11.1)*

5.



6.

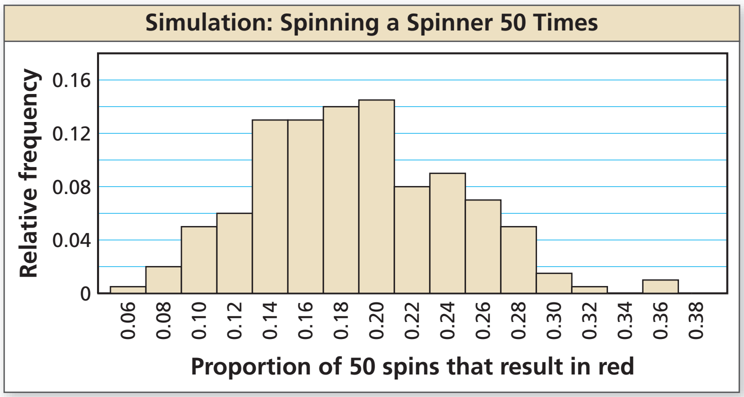


7. A survey of 1654 high school seniors determined that 1125 plan to attend college. Identify the population and the sample. Describe the sample. *(Section 11.2)*

8. A survey of all employees at a company found that the mean one-way daily commute to work of the employees is 25.5 minutes. Is the mean time a parameter or a statistic? Explain your reasoning. *(Section 11.2)*

9. A researcher records the number of bacteria present in several samples in a laboratory. Identify the method of data collection. *(Section 11.3)*

10. You spin a five-color spinner, which is divided into equal parts, five times and every time the spinner lands on red. You suspect the spinner favors red. The maker of the spinner claims that the spinner does not favor any color. You simulate spinning the spinner 50 times by repeatedly drawing 200 random samples of size 50. The histogram shows the results. Use the histogram to determine what you should conclude when you spin the actual spinner 50 times and the spinner lands on red (a) 9 times and (b) 19 times. *(Section 11.2)*



11. A local television station wants to find the number of hours per week people in the viewing area watch sporting events on television. The station surveys people at a nearby sports stadium. *(Section 11.3)*

a. Identify the type of sample described. **b.** Is the sample biased? Explain your reasoning.

c. Describe a method for selecting a random sample of 200 people to survey.

11.4 Experimental Design

Essential Question

How can you use an experiment to test a conjecture?

EXPLORATION 1 Using an Experiment

Work with a partner. Standard white playing dice are manufactured with black dots that are indentations, as shown. So, the side with six indentations is the lightest side and the side with one indentation is the heaviest side.

You make a conjecture that when you roll a standard playing die, the number 6 will come up more often than the number 1 because 6 is the lightest side. To test your conjecture, roll a standard playing die 25 times. Record the results in the table. Does the experiment confirm your conjecture? Explain your reasoning.

Number						
Rolls						

EXPLORATION 2 Analyzing an Experiment

Work with a partner. To overcome the imbalance of standard playing dice, one of the authors of this book invented and patented 12-sided dice, on which each number from 1 through 6 appears twice (on opposing sides). See *BigIdeasMath.com*.

As part of the patent process, a standard playing die was rolled 27,090 times. The results are shown below.

Number	1	2	3	4	5	6
Rolls	4293	4524	4492	4397	4623	4761

What can you conclude from the results of this experiment? Explain your reasoning.

CONSTRUCTING VIABLE ARGUMENTS

To be proficient in math, you need to make conjectures and perform experiments to explore the truth of your conjectures.

Communicate Your Answer

3. How can you use an experiment to test a conjecture?

4. Exploration 2 shows the results of rolling a standard playing die 27,090 times to test the conjecture in Exploration 1. Why do you think the number of trials was so large?

5. Make a conjecture about the outcomes of rolling the 12-sided die in Exploration 2. Then design an experiment that could be used to test your conjecture. Be sure that your experiment is practical to complete and includes enough trials to give meaningful results.

11.4 Lesson

What You Will Learn

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

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

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.

11.4 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **COMPLETE THE SENTENCE** Repetition of an experiment under the same or similar conditions is called __________.

2. **WRITING** Describe the difference between the control group and the treatment group in a controlled experiment.

Monitoring Progress and Modeling with Mathematics

In Exercises 3 and 4, 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. *(See Example 1.)*

3. ***Insomnia***

New Drug Improves Sleep

To test a new drug for insomnia, a pharmaceutical company randomly divided 200 adult volunteers into two groups. One group received the drug and one group received a placebo. After one month, the adults who took the drug slept 18% longer, while those who took the placebo experienced no significant change.

4. ***Dental Health***

Milk Fights Cavities

At a middle school, students can choose to drink milk or other beverages at lunch. Seventy-five students who chose milk were monitored for one year, as were 75 students who chose other beverages. At the end of the year, students in the "milk" group had 25% fewer cavities than students in the other group.

ERROR ANALYSIS In Exercises 5 and 6, describe and correct the error in describing the study.

A company's researchers want to study the effects of adding shea butter to their existing hair conditioner. They monitor the hair quality of 30 randomly selected customers using the regular conditioner and 30 randomly selected customers using the new shea butter conditioner.

5. ✗ The control group is individuals who do not use either of the conditioners.

6. ✗ The study is an observational study.

In Exercises 7–10, explain whether the research topic is best investigated through an experiment or an observational study. Then describe the design of the experiment or observational study. *(See Example 2.)*

7. A researcher wants to compare the body mass index of smokers and nonsmokers.

8. A restaurant chef wants to know which pasta sauce recipe is preferred by more diners.

9. A farmer wants to know whether a new fertilizer affects the weight of the fruit produced by strawberry plants.

10. You want to know whether homes that are close to parks or schools have higher property values.

11. **DRAWING CONCLUSIONS** A company wants to test whether a nutritional supplement has an adverse effect on an athlete's heart rate while exercising. Identify a potential problem, if any, with each experimental design. Then describe how you can improve it. *(See Example 3.)*

    a. The company randomly selects 250 athletes. Half of the athletes receive the supplement and their heart rates are monitored while they run on a treadmill. The other half of the athletes are given a placebo and their heart rates are monitored while they lift weights. The heart rates of the athletes who took the supplement significantly increased while exercising.

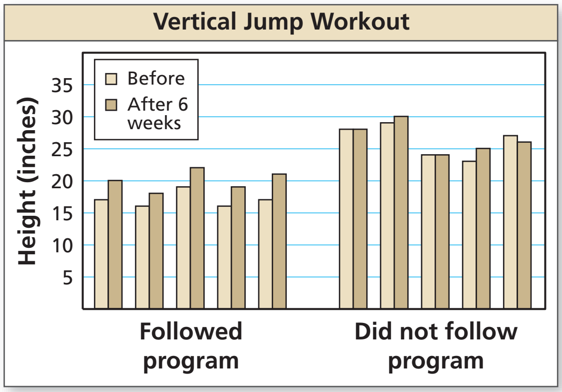
    b. The company selects 1000 athletes. The athletes are divided into two groups based on age. Within each age group, the athletes are randomly assigned to receive the supplement or the placebo. The athletes' heart rates are monitored while they run on a treadmill. There was no significant difference in the increases in heart rates between the two groups.

12. **DRAWING CONCLUSIONS** A researcher wants to test the effectiveness of reading novels on raising intelligence quotient (IQ) scores. Identify a potential problem, if any, with each experimental design. Then describe how you can improve it.

a. The researcher selects 500 adults and randomly divides them into two groups. One group reads novels daily and one group does not read novels. At the end of 1 year, each adult is evaluated and it is determined that neither group had an increase in IQ scores.

b. Fifty adults volunteer to spend time reading novels every day for 1 year. Fifty other adults volunteer to refrain from reading novels for 1 year. Each adult is evaluated and it is determined that the adults who read novels raised their IQ scores by 3 points more than the other group.

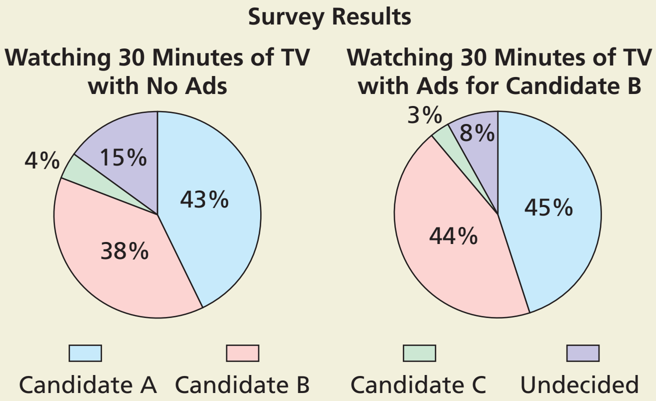
13. **DRAWING CONCLUSIONS** A fitness company claims that its workout program will increase vertical jump heights in 6 weeks. To test the workout program, 10 athletes are divided into two groups. The double bar graph shows the results of the experiment. Identify the potential problems with the experimental design. Then describe how you can improve it.



14. **WRITING** Explain why observational studies, rather than experiments, are usually used in astronomy.

15. **MAKING AN ARGUMENT** Your friend wants to determine whether the number of siblings has an effect on a student's grades. Your friend claims to be able to show causality between the number of siblings and grades. Is your friend correct? Explain.

16. **HOW DO YOU SEE IT?** To test the effect political advertisements have on voter preferences, a researcher selects 400 potential voters and randomly divides them into two groups. The circle graphs show the results of the study.



a. Is the study a randomized comparative experiment? Explain.

b. Describe the treatment.

c. Can you conclude that the political advertisements were effective? Explain.

17. **WRITING** Describe the *placebo effect* and how it affects the results of an experiment. Explain how a researcher can minimize the placebo effect.

18. **THOUGHT PROVOKING** Make a hypothesis about something that interests you. Design an experiment that could show that your hypothesis is probably true.

19. **REASONING** Will replicating an experiment on many individuals produce data that are more likely to accurately represent a population than performing the experiment only once? Explain.

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Draw a dot plot that represents the data. Identify the shape of the distribution. *(Skills Review Handbook)*

20. Ages: 24, 21, 22, 26, 22, 23, 25, 23, 23, 24, 20, 25

21. Golf strokes: 4, 3, 4, 3, 3, 2, 7, 5, 3, 4

Tell whether the function represents *exponential growth* or *exponential decay*. Then graph the function. *(Section 6.1)*

22. $y = 4^x$

23. $y = (0.95)^x$

24. $y = (0.2)^x$

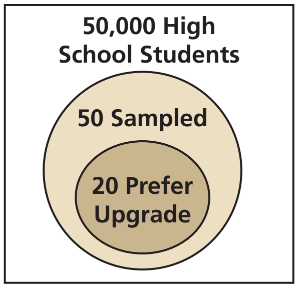
25. $y = (1.25)^x$

11.5 Making Inferences from Sample Surveys

Essential Question How can you use a sample survey to infer a conclusion about a population?

EXPLORATION 1 Making an Inference from a Sample

Work with a partner. You conduct a study to determine what percent of the high school students in your city would prefer an upgraded model of their current cell phone. Based on your intuition and talking with a few acquaintances, you think that 50% of high school students would prefer an upgrade. You survey 50 randomly chosen high school students and find that 20 of them prefer an upgraded model.



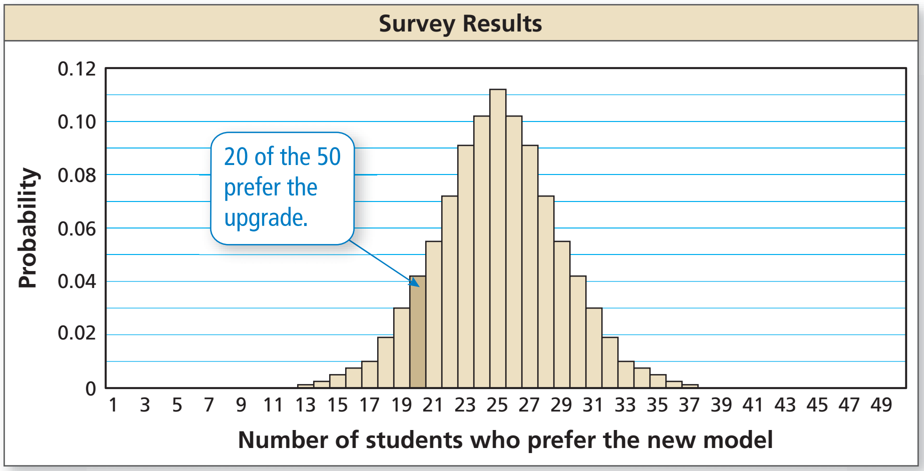
MODELING WITH MATHEMATICS

To be proficient in math, you need to apply the mathematics you know to solve problems arising in everyday life.

a. Based on your sample survey, what percent of the high school students in your city would prefer an upgraded model? Explain your reasoning.

b. In spite of your sample survey, is it still possible that 50% of the high school students in your city prefer an upgraded model? Explain your reasoning.

c. To investigate the likelihood that you could have selected a sample of 50 from a population in which 50% of the population does prefer an upgraded model, you create a binomial distribution as shown below. From the distribution, estimate the probability that exactly 20 students surveyed prefer an upgraded model. Is this event likely to occur? Explain your reasoning.



d. When making inferences from sample surveys, the sample must be random. In the situation described above, describe how you could design and conduct a survey using a random sample of 50 high school students who live in a large city.

Communicate Your Answer

2. How can you use a sample survey to infer a conclusion about a population?

3. In Exploration 1(c), what is the probability that exactly 25 students you survey prefer an upgraded model?

11.5 Lesson

Core Vocabulary

descriptive statistics, *p. 626*
inferential statistics, *p. 626*
margin of error, *p. 629*

Previous
statistic
parameter

What You Will Learn

- Estimate population parameters.
- Analyze estimated population parameters.
- Find margins of error for surveys.

Estimating Population Parameters

The study of statistics has two major branches: *descriptive statistics* and *inferential statistics*. **Descriptive statistics** involves the organization, summarization, and display of data. So far, you have been using descriptive statistics in your studies of data analysis and statistics. **Inferential statistics** involves using a sample to draw conclusions about a population. You can use statistics to make reasonable predictions, or *inferences*, about an entire population when the sample is representative of the population.

EXAMPLE 1 Estimating a Population Mean

The numbers of friends for a random sample of 40 teen users of a social networking website are shown in the table. Estimate the population mean μ.

Number of Friends				
281	342	229	384	320
247	298	248	312	445
385	286	314	260	186
287	342	225	308	343
262	220	320	310	150
274	291	300	410	255
279	351	370	257	350
369	215	325	338	278

REMEMBER

Recall that $\bar{x}$ denotes the sample mean. It is read as "*x* bar."

SOLUTION

To estimate the unknown population mean μ, find the sample mean $\bar{x}$.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{11{,}966}{40} = 299.15$$

So, the mean number of friends for all teen users of the website is about 299.

STUDY TIP

The probability that the population mean is *exactly* 299.15 is virtually 0, but the sample mean is a good estimate of μ.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

1. The data from another random sample of 30 teen users of the social networking website are shown in the table. Estimate the population mean μ.

Number of Friends				
305	237	261	374	341
257	243	352	330	189
297	418	275	288	307
295	288	341	322	271
209	164	363	228	390
313	315	263	299	285

Not every random sample results in the same estimate of a population parameter; there will be some sampling variability. Larger sample sizes, however, tend to produce more accurate estimates.

REMEMBER

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.

STUDY TIP

Statistics and probability provide information that you can use to weigh evidence and make decisions.

EXAMPLE 2 Estimating Population Proportions

A student newspaper wants to predict the winner of a city's mayoral election. Two candidates, A and B, are running for office. Eight staff members conduct surveys of randomly selected residents. The residents are asked whether they will vote for Candidate A. The results are shown in the table.

Sample Size	Number of Votes for Candidate A in the Sample	Percent of Votes for Candidate A in the Sample
5	2	40%
12	4	33.3%
20	12	60%
30	17	56.7%
50	29	58%
125	73	58.4%
150	88	58.7%
200	118	59%

a. Based on the results of the first two sample surveys, do you think Candidate A will win the election? Explain.

b. Based on the results in the table, do you think Candidate A will win the election? Explain.

SOLUTION

a. The results of the first two surveys (sizes 5 and 12) show that fewer than 50% of the residents will vote for Candidate A. Because there are only two candidates, one candidate needs more than 50% of the votes to win.

▶ Based on these surveys, you can predict Candidate A will not win the election.

b. As the sample sizes increase, the estimated percent of votes approaches 59%. You can predict that 59% of the city residents will vote for Candidate A.

▶ Because 59% of the votes are more than the 50% needed to win, you should feel confident that Candidate A will win the election.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

2. Two candidates are running for class president. The table shows the results of four surveys of random students in the class. The students were asked whether they will vote for the incumbent. Do you think the incumbent will be reelected? Explain.

Sample Size	Number of "Yes" Responses	Percent of Votes for Incumbent
10	7	70%
20	11	55%
30	13	43.3%
40	17	42.5%

Analyzing Estimated Population Parameters

An estimated population parameter is a hypothesis. You learned in Section 11.2 that one way to analyze a hypothesis is to perform a simulation.

EXAMPLE 3 Analyzing an Estimated Population Proportion

A national polling company claims 34% of U.S. adults say mathematics is the most valuable school subject in their lives. You survey a random sample of 50 adults.

a. What can you conclude about the accuracy of the claim that the population proportion is 0.34 when 15 adults in your survey say mathematics is the most valuable subject?

b. What can you conclude about the accuracy of the claim when 25 adults in your survey say mathematics is the most valuable subject?

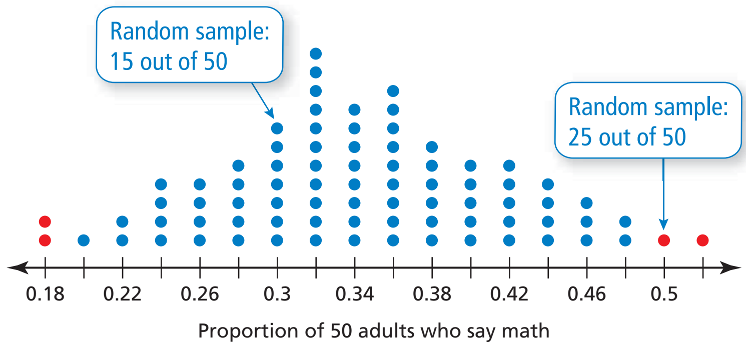
c. Assume that the true population proportion is 0.34. Estimate the variation among sample proportions using samples of size 50.

SOLUTION

```
randInt(0,99,50)
{76 10 27 54 41...
```

The polling company's claim (hypothesis) is that the population proportion of U.S. adults who say mathematics is the most valuable school subject is 0.34. To analyze this claim, simulate choosing 80 random samples of size 50 using a random number generator on a graphing calculator. Generate 50 random numbers from 0 to 99 for each sample. Let numbers 1 through 34 represent adults who say math. Find the sample proportions and make a dot plot showing the distribution of the sample proportions.

Simulation: Polling 50 Adults



STUDY TIP

The dot plot shows the results of one simulation. Results of other simulations may give slightly different results but the shape should be similar.

a. Note that 15 out of 50 corresponds to a sample proportion of $\frac{15}{50} = 0.3$. In the simulation, this result occurred in 7 of the 80 random samples. It is *likely* that 15 adults out of 50 would say math is the most valuable subject when the true population percentage is 34%. So, you can conclude the company's claim is probably accurate.

b. Note that 25 out of 50 corresponds to a sample proportion of $\frac{25}{50} = 0.5$. In the simulation, this result occurred in only 1 of the 80 random samples. So, it is *unlikely* that 25 adults out of 50 would say math is the most valuable subject when the true population percentage is 34%. So, you can conclude the company's claim is probably *not* accurate.

c. Note that the dot plot is fairly bell-shaped and symmetric, so the distribution is approximately normal. In a normal distribution, you know that about 95% of the possible sample proportions will lie within two standard deviations of 0.34. Excluding the two least and two greatest sample proportions, represented by red dots ● in the dot plot, leaves 76 of 80, or 95%, of the sample proportions. These 76 proportions range from 0.2 to 0.48. So, 95% of the time, a sample proportion should lie in the interval from 0.2 to 0.48.

INTERPRETING MATHEMATICAL RESULTS

Note that the sample proportion 0.3 in part (a) lies in this interval, while the sample proportion 0.5 in part (b) falls outside this interval.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

3. WHAT IF? In Example 3, what can you conclude about the accuracy of the claim that the population proportion is 0.34 when 21 adults in your random sample say mathematics is the most valuable subject?

Finding Margins of Error for Surveys

When conducting a survey, you need to make the size of your sample large enough so that it accurately represents the population. As the sample size increases, the *margin of error* decreases.

The **margin of error** gives a limit on how much the responses of the sample would differ from the responses of the population. For example, if 40% of the people in a poll favor a new tax law, and the margin of error is $\pm 4\%$, then it is likely that between 36% and 44% of the entire population favor a new tax law.

Core Concept

Margin of Error Formula

When a random sample of size n is taken from a large population, the margin of error is approximated by

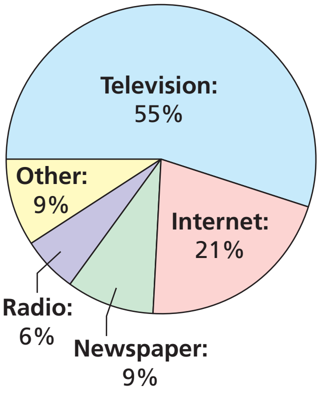
$$\text{Margin of error} = \pm \frac{1}{\sqrt{n}}.$$

This means that if the percent of the sample responding a certain way is p (expressed as a decimal), then the percent of the population who would respond the same way is likely to be between $p - \frac{1}{\sqrt{n}}$ and $p + \frac{1}{\sqrt{n}}$.

EXAMPLE 4 Finding a Margin of Error

In a survey of 2048 people in the U.S., 55% said that television is their main source of news. (a) What is the margin of error for the survey? (b) Give an interval that is likely to contain the exact percent of all people who use television as their main source of news.

Americans' Main News Source



SOLUTION

a. Use the margin of error formula.

$$\text{Margin of error} = \pm \frac{1}{\sqrt{n}} = \pm \frac{1}{\sqrt{2048}} \approx \pm 0.022$$

▶ The margin of error for the survey is about $\pm 2.2\%$.

b. To find the interval, subtract and add 2.2% to the percent of people surveyed who said television is their main source of news (55%).

$55\% - 2.2\% = 52.8\%$ $\qquad$ $55\% + 2.2\% = 57.2\%$

▶ It is likely that the exact percent of all people in the U.S. who use television as their main source of news is between 52.8% and 57.2%.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

4. In a survey of 1028 people in the U.S., 87% reported using the Internet. Give an interval that is likely to contain the exact percent of all people in the U.S. who use the Internet.

11.5 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **COMPLETE THE SENTENCE** The ____________ gives a limit on how much the responses of the sample would differ from the responses of the population.

2. **WRITING** What is the difference between descriptive and inferential statistics?

Monitoring Progress and Modeling with Mathematics

3. **PROBLEM SOLVING** The numbers of text messages sent each day by a random sample of 30 teen cellphone users are shown in the table. Estimate the population mean μ. *(See Example 1.)*

Number of Text Messages				
30	60	59	83	41
37	66	63	60	92
53	42	47	32	79
53	80	41	51	85
73	71	69	31	69
57	60	70	91	67

4. **PROBLEM SOLVING** The incomes for a random sample of 35 U.S. households are shown in the table. Estimate the population mean μ.

Income of U.S. Households				
14,300	52,100	74,800	51,000	91,500
72,800	50,500	15,000	37,600	22,100
40,000	65,400	50,000	81,100	99,800
43,300	32,500	76,300	83,400	24,600
30,800	62,100	32,800	21,900	64,400
73,100	20,000	49,700	71,000	45,900
53,200	45,500	55,300	19,100	63,100

5. **PROBLEM SOLVING** Use the data in Exercise 3 to answer each question.

   a. Estimate the population proportion ρ of teen cellphone users who send more than 70 text messages each day.

   b. Estimate the population proportion ρ of teen cellphone users who send fewer than 50 text messages each day.

6. **WRITING** A survey asks a random sample of U.S. teenagers how many hours of television they watch each night. The survey reveals that the sample mean is 3 hours per night. How confident are you that the average of all U.S. teenagers is exactly 3 hours per night? Explain your reasoning.

7. **DRAWING CONCLUSIONS** When the President of the United States vetoes a bill, the Congress can override the veto by a two-thirds majority vote in each House. Five news organizations conduct individual random surveys of U.S. Senators. The senators are asked whether they will vote to override the veto. The results are shown in the table. *(See Example 2.)*

Sample Size	Number of Votes to Override Veto	Percent of Votes to Override Veto
7	6	85.7%
22	16	72.7%
28	21	75%
31	17	54.8%
49	27	55.1%

   a. Based on the results of the first two surveys, do you think the Senate will vote to override the veto? Explain.

   b. Based on the results in the table, do you think the Senate will vote to override the veto? Explain.

8. **DRAWING CONCLUSIONS** Your teacher lets the students decide whether to have their test on Friday or Monday. The table shows the results from four surveys of randomly selected students in your grade who are taking the same class. The students are asked whether they want to have the test on Friday.

Sample Size	Number of "Yes" Responses	Percent of Votes
10	8	80%
20	12	60%
30	16	53.3%
40	18	45%

a. Based on the results of the first two surveys, do you think the test will be on Friday? Explain.

b. Based on the results in the table, do you think the test will be on Friday? Explain.

9. **MODELING WITH MATHEMATICS** A national polling company claims that 54% of U.S. adults are married. You survey a random sample of 50 adults. *(See Example 3.)*

a. What can you conclude about the accuracy of the claim that the population proportion is 0.54 when 31 adults in your survey are married?

b. What can you conclude about the accuracy of the claim that the population proportion is 0.54 when 19 adults in your survey are married?

c. Assume that the true population proportion is 0.54. Estimate the variation among sample proportions for samples of size 50.

10. **MODELING WITH MATHEMATICS** Employee engagement is the level of commitment and involvement an employee has toward the company and its values. A national polling company claims that only 29% of U.S. employees feel engaged at work. You survey a random sample of 50 U.S. employees.

a. What can you conclude about the accuracy of the claim that the population proportion is 0.29 when 16 employees feel engaged at work?

b. What can you conclude about the accuracy of the claim that the population proportion is 0.29 when 23 employees feel engaged at work?

c. Assume that the true population proportion is 0.29. Estimate the variation among sample proportions for samples of size 50.

In Exercises 11–16, find the margin of error for a survey that has the given sample size. Round your answer to the nearest tenth of a percent.

11. 260
12. 1000
13. 2024
14. 6400
15. 3275
16. 750

17. **ATTENDING TO PRECISION** In a survey of 1020 U.S. adults, 41% said that their top priority for saving is retirement. *(See Example 4.)*

a. What is the margin of error for the survey?

b. Give an interval that is likely to contain the exact percent of all U.S. adults whose top priority for saving is retirement.

18. **ATTENDING TO PRECISION** In a survey of 1022 U.S. adults, 76% said that more emphasis should be placed on producing domestic energy from solar power.

a. What is the margin of error for the survey?

b. Give an interval that is likely to contain the exact percent of all U.S. adults who think more emphasis should be placed on producing domestic energy from solar power.

19. **ERROR ANALYSIS** In a survey, 8% of adult Internet users said they participate in sports fantasy leagues online. The margin of error is ±4%. Describe and correct the error in calculating the sample size.

$$\pm 0.08 = \pm\frac{1}{\sqrt{n}}$$
$$0.0064 = \frac{1}{n}$$
$$n \approx 156$$

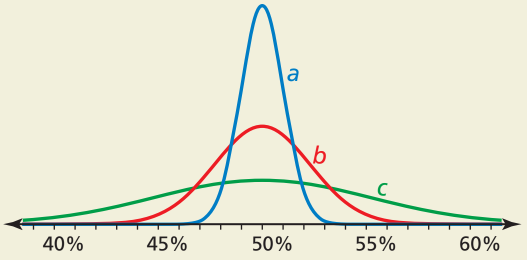
20. **ERROR ANALYSIS** In a random sample of 2500 consumers, 61% prefer Game A over Game B. Describe and correct the error in giving an interval that is likely to contain the exact percent of all consumers who prefer Game A over Game B.

Margin of error $= \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{2500}} = 0.02$

It is likely that the exact percent of all consumers who prefer Game A over Game B is between 60% and 62%.

21. **MAKING AN ARGUMENT** Your friend states that it is possible to have a margin of error between 0 and 100 percent, not including 0 or 100 percent. Is your friend correct? Explain your reasoning.

22. **HOW DO YOU SEE IT?** The figure shows the distribution of the sample proportions from three simulations using different sample sizes. Which simulation has the least margin of error? the greatest? Explain your reasoning.



23. **REASONING** A developer claims that the percent of city residents who favor building a new football stadium is likely between 52.3% and 61.7%. How many residents were surveyed?

24. **ABSTRACT REASONING** Suppose a random sample of size n is required to produce a margin of error of $\pm E$. Write an expression in terms of n for the sample size needed to reduce the margin of error to $\pm\frac{1}{2}E$. How many times must the sample size be increased to cut the margin of error in half? Explain.

25. **PROBLEM SOLVING** A survey reported that 47% of the voters surveyed, or about 235 voters, said they voted for Candidate A and the remainder said they voted for Candidate B.

a. How many voters were surveyed?

b. What is the margin of error for the survey?

c. For each candidate, find an interval that is likely to contain the exact percent of all voters who voted for the candidate.

d. Based on your intervals in part (c), can you be confident that Candidate B won? If not, how many people in the sample would need to vote for Candidate B for you to be confident that Candidate B won? (*Hint:* Find the least number of voters for Candidate B so that the intervals do not overlap.)

26. **THOUGHT PROVOKING** Consider a large population in which ρ percent (in decimal form) have a certain characteristic. To be reasonably sure that you are choosing a sample that is representative of a population, you should choose a random sample of n people where

$$n > 9\left(\frac{1-\rho}{\rho}\right).$$

a. Suppose $\rho = 0.5$. How large does n need to be?

b. Suppose $\rho = 0.01$. How large does n need to be?

c. What can you conclude from parts (a) and (b)?

27. **CRITICAL THINKING** In a survey, 52% of the respondents said they prefer sports drink X and 48% said they prefer sports drink Y. How many people would have to be surveyed for you to be confident that sports drink X is truly preferred by more than half the population? Explain.

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Find the inverse of the function. *(Section 6.3)*

28. $y = 10^{x-3}$

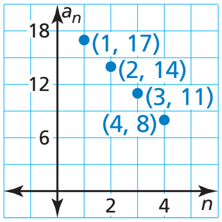
29. $y = 2^x - 5$

30. $y = \ln(x + 5)$

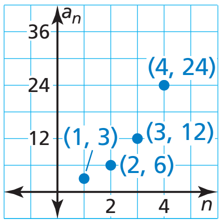
31. $y = \log_6 x - 1$

Determine whether the graph represents an arithmetic sequence or a geometric sequence. Then write a rule for the nth term. *(Section 8.2 and Section 8.3)*

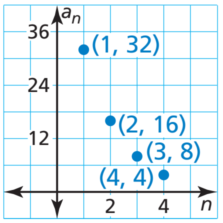
32.



33.



34.



11.6 Making Inferences from Experiments

Essential Question

How can you test a hypothesis about an experiment?

EXPLORATION 1 Resampling Data

Work with a partner. A randomized comparative experiment tests whether water with dissolved calcium affects the yields of yellow squash plants. The table shows the results.

Yield (kilograms)	
Control Group	Treatment Group
1.0	1.1
1.2	1.3
1.5	1.4
0.9	1.2
1.1	1.0
1.4	1.7
0.8	1.8
0.9	1.1
1.3	1.1
1.6	1.8

a. Find the mean yield of the control group and the mean yield of the treatment group. Then find the difference of the two means. Record the results.

b. Write each yield measurement from the table on an equal-sized piece of paper. Place the pieces of paper in a bag, shake, and randomly choose 10 pieces of paper. Call this the "control" group, and call the 10 pieces in the bag the "treatment" group. Then repeat part (a) and return the pieces to the bag. Perform this resampling experiment five times.

c. How does the difference in the means of the control and treatment groups compare with the differences resulting from chance?

EXPLORATION 2 Evaluating Results

Work as a class. To conclude that the treatment is responsible for the difference in yield, you need strong evidence to reject the hypothesis:

Water dissolved in calcium has no effect on the yields of yellow squash plants.

To evaluate this hypothesis, compare the experimental difference of means with the resampling differences.

a. Collect all the resampling differences of means found in Exploration 1(b) for the whole class and display these values in a histogram.

b. Draw a vertical line on your class histogram to represent the experimental difference of means found in Exploration 1(a).

c. Where on the histogram should the experimental difference of means lie to give evidence for rejecting the hypothesis?

d. Is your class able to reject the hypothesis? Explain your reasoning.

MODELING WITH MATHEMATICS

To be proficient in math, you need to identify important quantities in a practical situation, map their relationships using such tools as diagrams and graphs, and analyze those relationships mathematically to draw conclusions.

Communicate Your Answer

3. How can you test a hypothesis about an experiment?

4. The randomized comparative experiment described in Exploration 1 is replicated and the results are shown in the table. Repeat Explorations 1 and 2 using this data set. Explain any differences in your answers.

	Yield (kilograms)									
Control Group	0.9	0.9	1.4	0.6	1.0	1.1	0.7	0.6	1.2	1.3
Treatment Group	1.0	1.2	1.2	1.3	1.0	1.8	1.7	1.2	1.0	1.9

11.6 Lesson

Core Vocabulary

Previous
randomized comparative experiment
control group
treatment group
mean
dot plot
outlier
simulation
hypothesis

What You Will Learn

- Organize data from an experiment with two samples.
- Resample data using a simulation to analyze a hypothesis.
- Make inferences about a treatment.

Experiments with Two Samples

In this lesson, you will compare data from two samples in an experiment to make inferences about a treatment using a method called *resampling*. Before learning about this method, consider the experiment described in Example 1.

EXAMPLE 1 Organizing Data from an Experiment

A randomized comparative experiment tests whether a soil supplement affects the total yield (in kilograms) of cherry tomato plants. The control group has 10 plants and the treatment group, which receives the soil supplement, has 10 plants. The table shows the results.

	Total Yield of Tomato Plants (kilograms)									
Control Group	1.2	1.3	0.9	1.4	2.0	1.2	0.7	1.9	1.4	1.7
Treatment Group	1.4	0.9	1.5	1.8	1.6	1.8	2.4	1.9	1.9	1.7

a. Find the mean yield of the control group, $\overline{x}_{\text{control}}$.

b. Find the mean yield of the treatment group, $\overline{x}_{\text{treatment}}$.

c. Find the experimental difference of the means, $\overline{x}_{\text{treatment}} - \overline{x}_{\text{control}}$.

d. Display the data in a double dot plot.

e. What can you conclude?

SOLUTION

a. $\overline{x}_{\text{control}} = \dfrac{1.2 + 1.3 + 0.9 + 1.4 + 2.0 + 1.2 + 0.7 + 1.9 + 1.4 + 1.7}{10} = \dfrac{13.7}{10} = 1.37$

- The mean yield of the control group is 1.37 kilograms.

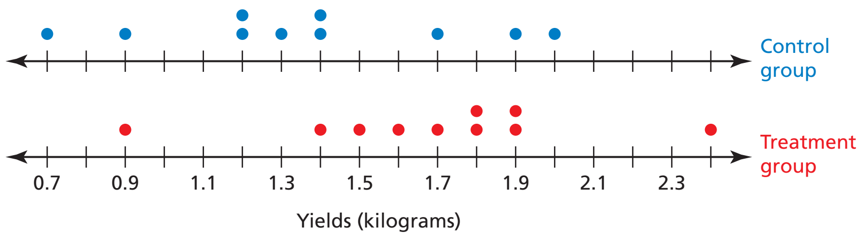
b. $\overline{x}_{\text{treatment}} = \dfrac{1.4 + 0.9 + 1.5 + 1.8 + 1.6 + 1.8 + 2.4 + 1.9 + 1.9 + 1.7}{10} = \dfrac{16.9}{10} = 1.69$

- The mean yield of the treatment group is 1.69 kilograms.

c. $\overline{x}_{\text{treatment}} - \overline{x}_{\text{control}} = 1.69 - 1.37 = 0.32$

- The experimental difference of the means is 0.32 kilogram.

d.



e. The plot of the data shows that the two data sets tend to be fairly symmetric and have no extreme values (outliers). So, the mean is a suitable measure of center. The mean yield of the treatment group is 0.32 kilogram more than the control group. It appears that the soil supplement might be slightly effective, but the sample size is small and the difference could be due to chance.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

1. In Example 1, interpret the meaning of $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$ when the difference is (a) negative, (b) zero, and (c) positive.

Resampling Data Using a Simulation

The samples in Example 1 are too small to make inferences about the treatment. Statisticians have developed a method called resampling to overcome this problem. Here is one way to resample: combine the measurements from both groups, and repeatedly create new "control" and "treatment" groups at random from the measurements without repeats. Example 2 shows one resampling of the data in Example 1.

EXAMPLE 2 Resampling Data Using a Simulation

Resample the data in Example 1 using a simulation. Use the mean yields of the new control and treatment groups to calculate the difference of the means.

SOLUTION

Step 1 Combine the measurements from both groups and assign a number to each value. Let the numbers 1 through 10 represent the data in the original control group, and let the numbers 11 through 20 represent the data in the original treatment group, as shown.

original control group	1.2	1.3	0.9	1.4	2.0	1.2	0.7	1.9	1.4	1.7
assigned number	1	2	3	4	5	6	7	8	9	10

original treatment group	1.4	0.9	1.5	1.8	1.6	1.8	2.4	1.9	1.9	1.7
assigned number	11	12	13	14	15	16	17	18	19	20

Step 2 Use a random number generator. Randomly generate 20 numbers from 1 through 20 *without repeating a number*. The table shows the results.

```
randIntNoRep(1,20)
{14 19 4 3 18 9...
```

14	19	4	3	18	9	5	15	2	7
1	17	20	16	6	8	13	12	11	10

Use the first 10 numbers to make the new control group, and the next 10 to make the new treatment group. The results are shown in the next table.

	Resample of Tomato Plant Yields (kilograms)									
New Control Group	1.8	1.9	1.4	0.9	1.9	1.4	2.0	1.6	1.3	0.7
New Treatment Group	1.2	2.4	1.7	1.8	1.2	1.9	1.5	0.9	1.4	1.7

Step 3 Find the mean yields of the new control and treatment groups.

$$\bar{x}_{\text{new control}} = \frac{1.8 + 1.9 + 1.4 + 0.9 + 1.9 + 1.4 + 2.0 + 1.6 + 1.3 + 0.7}{10} = \frac{14.9}{10} = 1.49$$

$$\bar{x}_{\text{new treatment}} = \frac{1.2 + 2.4 + 1.7 + 1.8 + 1.2 + 1.9 + 1.5 + 0.9 + 1.4 + 1.7}{10} = \frac{15.7}{10} = 1.57$$

So, $\bar{x}_{\text{new treatment}} - \bar{x}_{\text{new control}} = 1.57 - 1.49 = 0.08$. This is less than the experimental difference found in Example 1.

Making Inferences About a Treatment

To perform an analysis of the data in Example 1, you will need to resample the data more than once. After resampling many times, you can see how often you get differences between the new groups that are at least as large as the one you measured.

EXAMPLE 3 Making Inferences About a Treatment

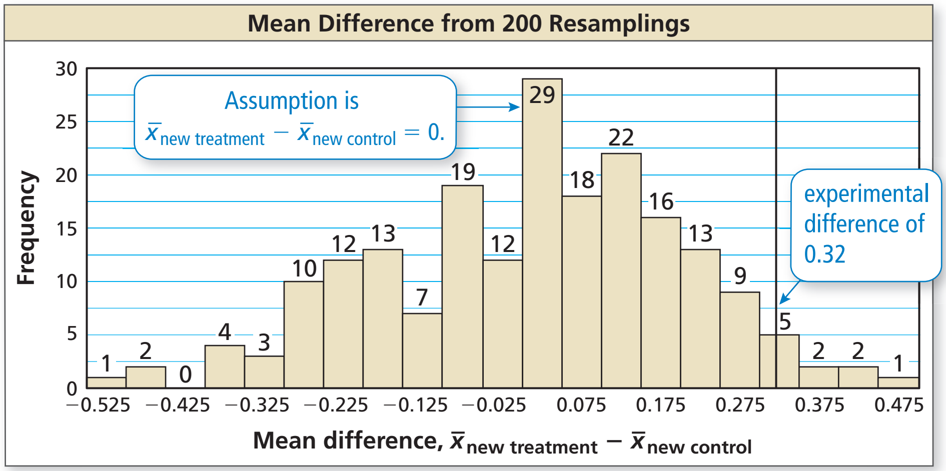
To conclude that the treatment in Example 1 is responsible for the difference in yield, you need to analyze this hypothesis:

The soil nutrient has no effect on the yield of the cherry tomato plants.

Simulate 200 resamplings of the data in Example 1. Compare the experimental difference of 0.32 from Example 1 with the resampling differences. What can you conclude about the hypothesis? Does the soil nutrient have an effect on the yield?

SOLUTION

The histogram shows the results of the simulation. The histogram is approximately bell-shaped and fairly symmetric, so the differences have an approximately normal distribution.



Note that the hypothesis assumes that the difference of the mean yields is 0. The experimental difference of 0.32, however, lies close to the right tail. From the graph, there are about 5 to 10 values out of 200 that are greater than 0.32, which is at most 5% of the values. Also, the experimental difference falls outside the middle 90% of the resampling differences. (The middle 90% is the area of the bars from −0.275 to 0.275, which contains 180 of the 200 values, or 90%.) This means it is unlikely to get a difference this large when you assume that the difference is 0, suggesting the control group and the treatment group differ.

INTERPRETING MATHEMATICAL RESULTS

With this conclusion, you can be 90% confident that the soil supplement does have an effect.

▶ You can conclude that the hypothesis is most likely false. So, the soil nutrient *does* have an effect on the yield of cherry tomato plants. Because the mean difference is positive, the treatment *increases* the yield.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

2. In Example 3, what are the consequences of concluding that the hypothesis is false when it is actually true?

11.6 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **COMPLETE THE SENTENCE** A method in which new samples are repeatedly drawn from the data set is called ____________.

2. **DIFFERENT WORDS, SAME QUESTION** Which is different? Find "both" answers.

What is the experimental difference of the means?

What is $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$?

	Weight of Tumor (grams)					
Control Group	3.3	3.2	3.7	3.5	3.3	3.4
Treatment Group	0.4	0.6	0.5	0.6	0.7	0.5

What is the square root of the average of the squared differences from −2.85?

What is the difference between the mean of the treatment group and the mean of the control group?

Monitoring Progress and Modeling with Mathematics

3. **PROBLEM SOLVING** A randomized comparative experiment tests whether music therapy affects the depression scores of college students. The depression scores range from 20 to 80, with scores greater than 50 being associated with depression. The control group has eight students and the treatment group, which receives the music therapy, has eight students. The table shows the results. *(See Example 1.)*

	Depression Score							
Control Group	49	45	43	47	46	45	47	46
Treatment Group	39	40	39	37	41	40	42	43

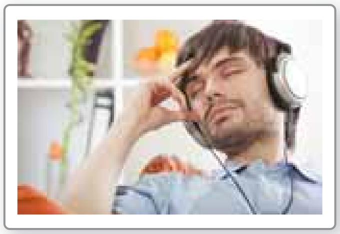
a. Find the mean score of the control group.

b. Find the mean score of the treatment group.

c. Find the experimental difference of the means.

d. Display the data in a double dot plot.

e. What can you conclude?



4. **PROBLEM SOLVING** A randomized comparative experiment tests whether low-level laser therapy affects the waist circumference of adults. The control group has eight adults and the treatment group, which receives the low-level laser therapy, has eight adults. The table shows the results.

	Circumference (inches)							
Control Group	34.6	35.4	33	34.6	35.2	35.2	36.2	35
Treatment Group	31.4	33	32.4	32.6	33.4	33.4	34.8	33

a. Find the mean circumference of the control group.

b. Find the mean circumference of the treatment group.

c. Find the experimental difference of the means.

d. Display the data in a double dot plot.

e. What can you conclude?

5. **ERROR ANALYSIS** In a randomized comparative experiment, the mean score of the treatment group is 11 and the mean score of the control group is 16. Describe and correct the error in interpreting the experimental difference of the means.

✗ $\bar{x}_{\text{control}} - \bar{x}_{\text{treatment}} = 16 - 11 = 5$

So, you can conclude the treatment increases the score.

6. **REASONING** In Exercise 4, interpret the meaning of $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$ when the difference is positive, negative, and zero.

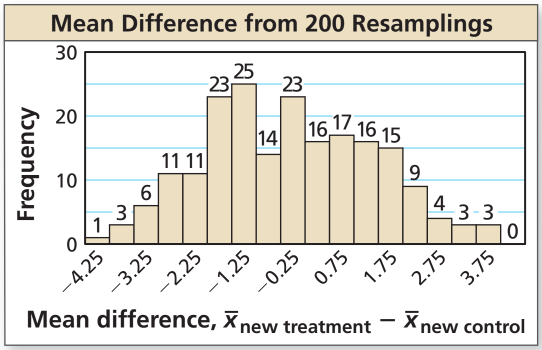
7. **MODELING WITH MATHEMATICS** Resample the data in Exercise 3 using a simulation. Use the means of the new control and treatment groups to calculate the difference of the means. *(See Example 2.)*

8. **MODELING WITH MATHEMATICS** Resample the data in Exercise 4 using a simulation. Use the means of the new control and treatment groups to calculate the difference of the means.

9. **DRAWING CONCLUSIONS** To analyze the hypothesis below, use the histogram which shows the results from 200 resamplings of the data in Exercise 3.

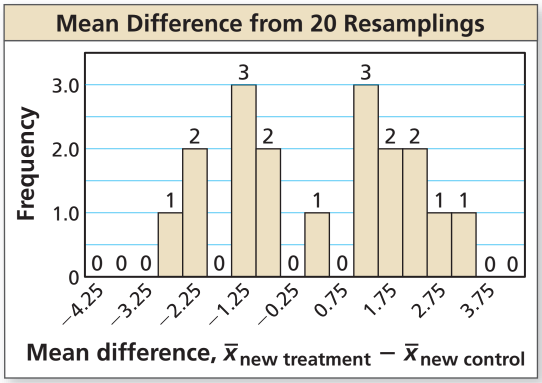
   Music therapy has no effect on the depression score.

   Compare the experimental difference in Exercise 3 with the resampling differences. What can you conclude about the hypothesis? Does music therapy have an effect on the depression score? *(See Example 3.)*

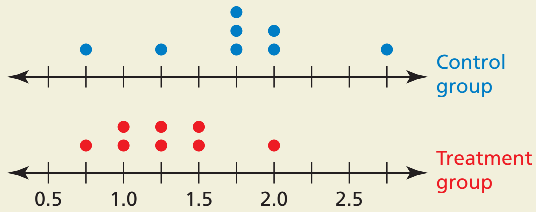


10. **DRAWING CONCLUSIONS** Suppose the experimental difference of the means in Exercise 3 had been −0.75. Compare this experimental difference of means with the resampling differences in the histogram in Exercise 9. What can you conclude about the hypothesis? Does music therapy have an effect on the depression score?

11. **WRITING** Compare the histogram in Exercise 9 to the histogram below. Determine which one provides stronger evidence against the hypothesis, *Music therapy has no effect on the depression score*. Explain.



12. **HOW DO YOU SEE IT?** Without calculating, determine whether the experimental difference, $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$, is positive, negative, or zero. What can you conclude about the effect of the treatment? Explain.



13. **MAKING AN ARGUMENT** Your friend states that the mean of the resampling differences of the means should be close to 0 as the number of resamplings increase. Is your friend correct? Explain your reasoning.

14. **THOUGHT PROVOKING** Describe an example of an observation that can be made from an experiment. Then give four possible inferences that could be made from the observation.

15. **CRITICAL THINKING** In Exercise 4, how many resamplings of the treatment and control groups are theoretically possible? Explain.

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Factor the polynomial completely. *(Section 4.4)*

16. $5x^3 - 15x^2$ 17. $y^3 - 8$ 18. $z^3 + 5z^2 - 9z - 45$ 19. $81w^4 - 16$

Determine whether the inverse of *f* is a function. Then find the inverse. *(Section 7.5)*

20. $f(x) = \dfrac{3}{x + 5}$ 21. $f(x) = \dfrac{1}{2x - 1}$ 22. $f(x) = \dfrac{2}{x} - 4$ 23. $f(x) = \dfrac{3}{x^2} + 1$

11.4–11.6 What Did You Learn?

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*
descriptive statistics, *p. 626*
inferential statistics, *p. 626*
margin of error, *p. 629*

Core Concepts

Section 11.4

Randomization in Experiments and Observational Studies, *p. 621*
Comparative Studies and Causality, *p. 621*
Analyzing Experimental Designs, *p. 622*

Section 11.5

Estimating Population Parameters, *p. 626*
Analyzing Estimated Population Parameters, *p. 628*

Section 11.6

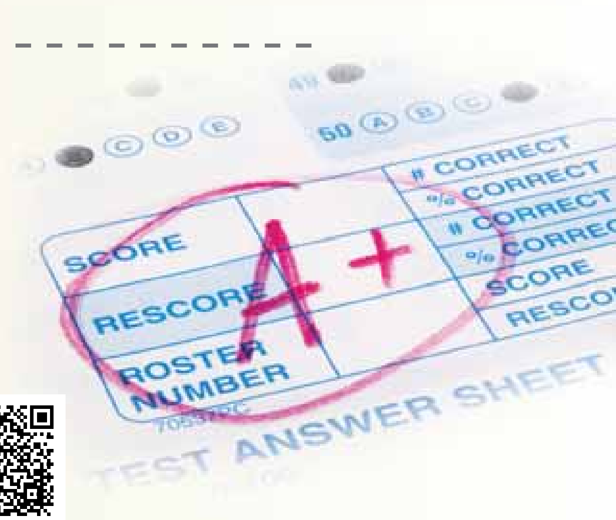
Experiments with Two Samples, *p. 634*
Resampling Data Using Simulations, *p. 635*
Making Inferences About Treatments, *p. 636*

Mathematical Practices

1. In Exercise 7 on page 623, find a partner and discuss your answers. What questions should you ask your partner to determine whether an observational study or an experiment is more appropriate?
2. In Exercise 23 on page 632, how did you use the given interval to find the sample size?

Performance Task

Curving the Test



Test scores are sometimes curved for different reasons using different techniques. Curving began with the assumption that a good test would result in scores that were normally distributed about a C average. Is this assumption valid? Are test scores in your class normally distributed? If not, how are they distributed? Which curving algorithms preserve the distribution and which algorithms change it?

To explore the answers to these questions and more, go to ***BigIdeasMath.com***.

11 Chapter Review

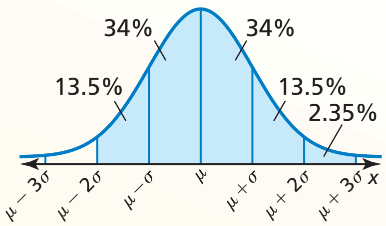
Dynamic Solutions available at *BigIdeasMath.com*

11.1 Using Normal Distributions

(pp. 595–602)

A normal distribution has mean μ and standard deviation σ. An x-value is randomly selected from the distribution. Find $P(\mu - 2\sigma \le x \le \mu + 3\sigma)$.

The probability that a randomly selected x-value lies between $\mu - 2\sigma$ and $\mu + 3\sigma$ is the shaded area under the normal curve shown.



$P(\mu - 2\sigma \le x \le \mu + 3\sigma) = 0.135 + 0.34 + 0.34 + 0.135 + 0.0235 = 0.9735$

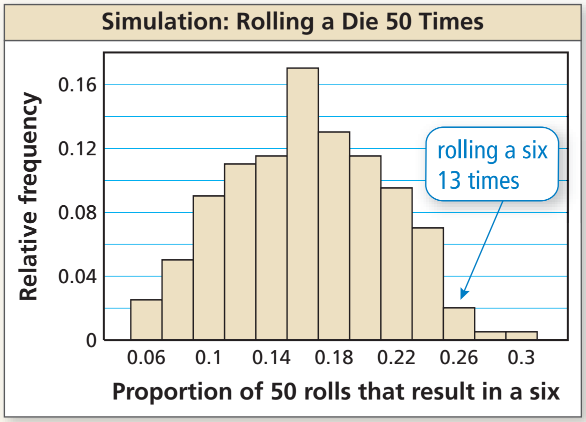
1. A normal distribution has mean μ and standard deviation σ. An x-value is randomly selected from the distribution. Find $P(x \le \mu - 3\sigma)$.
2. The scores received by juniors on the math portion of the PSAT are normally distributed with a mean of 48.6 and a standard deviation of 11.4. What is the probability that a randomly selected score is at least 76?

11.2 Populations, Samples, and Hypotheses

(pp. 603–608)

You suspect a die favors the number six. The die maker claims the die does not favor any number. What should you conclude when you roll the actual die 50 times and get a six 13 times?

The maker's claim, or hypothesis, is "the die does not favor any number." This is the same as saying that the proportion of sixes rolled, in the long run, is $\frac{1}{6}$. So, assume the probability of rolling a six is $\frac{1}{6}$. Simulate the rolling of the die by repeatedly drawing 200 random samples of size 50 from a population of numbers from one through six. Make a histogram of the distribution of the sample proportions.



Getting a six 13 times corresponds to a proportion of $\frac{13}{50} = 0.26$. In the simulation, this result had a relative frequency of 0.02. Because this result is unlikely to occur by chance, you can conclude that the maker's claim is most likely false.

3. To estimate the average number of miles driven by U.S. motorists each year, a researcher conducts a survey of 1000 drivers, records the number of miles they drive in a year, and then determines the average. Identify the population and the sample.
4. A pitcher throws 40 fastballs in a game. A baseball analyst records the speeds of 10 fastballs and finds that the mean speed is 92.4 miles per hour. Is the mean speed a parameter or a statistic? Explain.
5. A prize on a game show is placed behind either Door A or Door B. You suspect the prize is more often behind Door A. The show host claims the prize is randomly placed behind either door. What should you conclude when the prize is behind Door A for 32 out of 50 contestants?

11.3 Collecting Data (pp. 609–616)

You want to determine how many people in the senior class plan to study mathematics after high school. You survey every senior in your calculus class. Identify the type of sample described and determine whether the sample is biased.

▶ You select students who are readily available. So, the sample is a *convenience* sample. The sample is biased because students in a calculus class are more likely to study mathematics after high school.

6. A researcher wants to determine how many people in a city support the construction of a new road connecting the high school to the north side of the city. Fifty residents from each side of the city are surveyed. Identify the type of sample described and determine whether the sample is biased.

7. A researcher records the number of people who use a coupon when they dine at a certain restaurant. Identify the method of data collection.

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 think the city should replace the outdated police cars it is using?"

11.4 Experimental Design (pp. 619–624)

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.

Headphones Hurt Hearing
A study of 100 college and high school students compared their times spent listening to music using headphones with hearing loss. Twelve percent of people who listened to headphones more than one hour per day were found to have measurable hearing loss over the course of the three-year study.

▶ 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 headphone use impairs hearing ability may or may not be valid. For instance, people who listen to more than an hour of music per day may be more likely to attend loud concerts that are known to affect hearing.

9. A restaurant manager wants to know which type of sandwich bread attracts the most repeat customers. Is the topic best investigated through an experiment or an observational study? Describe how you would design the experiment or observational study.

10. A researcher wants to test the effectiveness of a sleeping pill. Identify a potential problem, if any, with the experimental design below. Then describe how you can improve it.

The researcher asks for 16 volunteers who have insomnia. Eight volunteers are given the sleeping pill and the other 8 volunteers are given a placebo. Results are recorded for 1 month.

11. 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.

Cleaner Cars in Less Time!
To test the new design of a car wash, an engineer gathered 80 customers and randomly divided them into two groups. One group used the old design to wash their cars and one group used the new design to wash their cars. Users of the new car wash design were able to wash their cars 30% faster.

11.5 Making Inferences from Sample Surveys (pp. 625–632)

Before the Thanksgiving holiday, in a survey of 2368 people, 85% said they are thankful for the health of their family. What is the margin of error for the survey?

Use the margin of error formula.

$$\text{Margin of error} = \pm\frac{1}{\sqrt{n}} = \pm\frac{1}{\sqrt{2368}} \approx \pm 0.021$$

▶ The margin of error for the survey is about ±2.1%.

12. In a survey of 1017 U.S. adults, 62% said that they prefer saving money over spending it. Give an interval that is likely to contain the exact percent of all U.S. adults who prefer saving money over spending it.

13. There are two candidates for homecoming king. The table shows the results from four random surveys of the students in the school. The students were asked whether they will vote for Candidate A. Do you think Candidate A will be the homecoming king? Explain.

Sample Size	Number of "Yes" Responses	Percent of Votes
8	6	75%
22	14	63.6%
34	16	47.1%
62	29	46.8%

11.6 Making Inferences from Experiments (pp. 633–638)

A randomized comparative experiment tests whether a new fertilizer affects the length (in inches) of grass after one week. The control group has 10 sections of land and the treatment group, which is fertilized, has 10 sections of land. The table shows the results.

	Grass Length (inches)									
Control Group	4.5	4.5	4.8	4.4	4.4	4.7	4.3	4.5	4.1	4.2
Treatment Group	4.6	4.8	5.0	4.8	4.7	4.6	4.9	4.9	4.8	4.4

a. Find the experimental difference of the means, $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$.

$\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}} = 4.75 - 4.44 = 0.31$

▶ The experimental difference of the means is 0.31 inch.

b. What can you conclude?

▶ The two data sets tend to be fairly symmetric and have no extreme values. So, the mean is a suitable measure of center. The mean length of the treatment group is 0.31 inch longer than the control group. It appears that the fertilizer might be slightly effective, but the sample size is small and the difference could be due to chance.

14. Describe how to use a simulation to resample the data in the example above. Explain how this allows you to make inferences about the data when the sample size is small.

11 Chapter Test

1. Market researchers want to know whether more men or women buy their product. Explain whether this research topic is best investigated through an experiment or an observational study. Then describe the design of the experiment or observational study.

2. You want to survey 100 of the 2774 four-year colleges in the United States about their tuition cost. Describe a method for selecting a random sample of colleges to survey.

3. The grade point averages of all the students in a high school are normally distributed with a mean of 2.95 and a standard deviation of 0.72. Are these numerical values parameters or statistics? Explain.

A normal distribution has a mean of 72 and a standard deviation of 5. Find the probability that a randomly selected x-value from the distribution is in the given interval.

4. between 67 and 77 **5.** at least 75 **6.** at most 82

7. A researcher wants to test the effectiveness of a new medication designed to lower blood pressure. Identify a potential problem, if any, with the experimental design. Then describe how you can improve it.

The researcher identifies 30 people with high blood pressure. Fifteen people with the highest blood pressures are given the medication and the other 15 are given a placebo. After 1 month, the subjects are evaluated.

8. A randomized comparative experiment tests whether a vitamin supplement increases human bone density (in grams per square centimeter). The control group has eight people and the treatment group, which receives the vitamin supplement, has eight people. The table shows the results.

	Bone Density (g/cm^2)							
Control Group	0.9	1.2	1.0	0.8	1.3	1.1	0.9	1.0
Treatment Group	1.2	1.0	0.9	1.3	1.2	0.9	1.3	1.2

a. Find the mean yields of the control group, $\overline{x}_{\text{control}}$, and the treatment group, $\overline{x}_{\text{treatment}}$.

b. Find the experimental difference of the means, $\overline{x}_{\text{treatment}} - \overline{x}_{\text{control}}$.

c. Display the data in a double dot plot. What can you conclude?

d. Five hundred resamplings of the data are simulated. Out of the 500 resampling differences, 231 are greater than the experimental difference in part (b). What can you conclude about the hypothesis, *The vitamin supplement has no effect on human bone density*? Explain your reasoning.

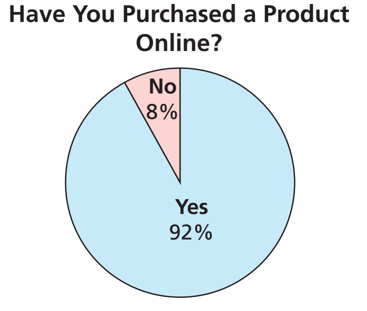
9. In a recent survey of 1600 randomly selected U.S. adults, 81% said they have purchased a product online.

a. Identify the population and the sample. Describe the sample.

b. Find the margin of error for the survey.

c. Give an interval that is likely to contain the exact percent of all U.S. adults who have purchased a product online.

d. You survey 75 teachers at your school. The results are shown in the graph. Would you use the recent survey or your survey to estimate the percent of U.S. adults who have purchased a product online? Explain.



11 Cumulative Assessment

1. Your friend claims any system formed by three of the following equations will have exactly one solution.

$3x + y + 3z = 6$	$x + y + z = 2$	$4x - 2y + 4z = 8$
$x - y + z = 2$	$2x + y + z = 4$	$3x + y + 9z = 12$

   a. Write a linear system that would support your friend's claim.

   b. Write a linear system that shows your friend's claim is incorrect.

2. Which of the following samples are biased? If the sample is biased, explain why it is biased.

   Ⓐ A restaurant asks customers to participate in a survey about the food sold at the restaurant. The restaurant uses the surveys that are returned.

   Ⓑ You want to know the favorite sport of students at your school. You randomly select athletes to survey at the winter sports banquet.

   Ⓒ The owner of a store wants to know whether the store should stay open 1 hour later each night. Each cashier surveys every fifth customer.

   Ⓓ The owner of a movie theater wants to know whether the volume of its movies is too loud. Patrons under the age of 18 are randomly surveyed.

3. A survey asks adults about their favorite way to eat ice cream. The results of the survey are displayed in the table shown.

Survey Results	
Cup	45%
Cone	29%
Sundae	18%
Other	8%
(*margin of error* ±2.11%)	

   a. How many people were surveyed?

   b. Why might the conclusion, "Adults generally do not prefer to eat their ice cream in a cone" be inaccurate to draw from this data?

   c. You decide to test the results of the poll by surveying adults chosen at random. What is the probability that at least three out of the six people you survey prefer to eat ice cream in a cone?

   d. Four of the six respondents in your study said they prefer to eat their ice cream in a cone. You conclude that the other survey is inaccurate. Why might this conclusion be incorrect?

   e. What is the margin of error for your survey?

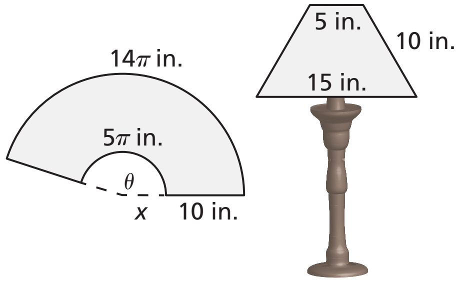
4. You are making a lampshade out of fabric for the lamp shown. The pattern for the lampshade is shown in the diagram on the left.

a. Use the smaller sector to write an equation that relates θ and x.

b. Use the larger sector to write an equation that relates θ and $x + 10$.

c. Solve the system of equations from parts (a) and (b) for x and θ.

d. Find the amount of fabric (in square inches) that you will use to make the lampshade.

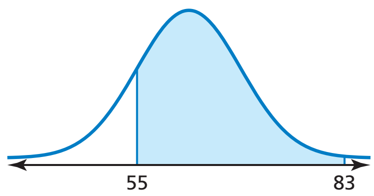


5. For all students taking the Medical College Admission Test over a period of 3 years, the mean score was 25.1. During the same 3 years, a group of 1000 students who took the test had a mean score of 25.3. Classify each mean as a parameter or a statistic. Explain.

6. Complete the table for the four equations. Explain your reasoning.

Equation	Is the inverse a function?		Is the function its own inverse?	
	Yes	No	Yes	No
$y = -x$				
$y = 3 \ln x + 2$				
$y = \left(\frac{1}{x}\right)^2$				
$y = \frac{x}{x - 1}$				

7. The normal distribution shown has mean 63 and standard deviation 8. Find the percent of the area under the normal curve that is represented by the shaded region. Then describe another interval under the normal curve that has the same area.



8. Which of the rational expressions *cannot* be simplified?

Ⓐ $\dfrac{2x^2 + 5x - 3}{x^2 - 7x + 12}$

Ⓑ $\dfrac{3x^3 + 21x^2 + 30x}{x^2 - 25}$

Ⓒ $\dfrac{x^3 + 27}{x^2 - 3x + 9}$

Ⓓ $\dfrac{x^3 + 2x^2 - 8x - 16}{2x^2 - 21x + 55}$