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?

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.

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

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

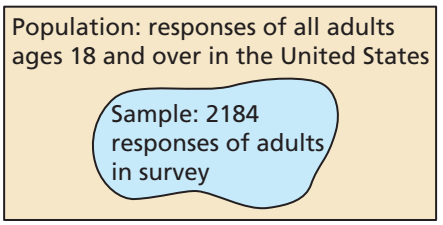
Identify the population and the sample. Describe the sample.

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

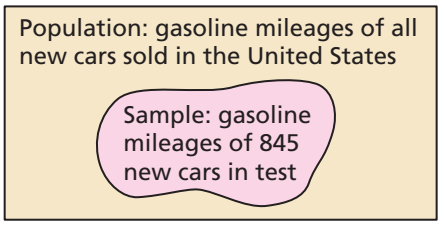
b. To estimate the gasoline mileage of new cars sold in the United States, a consumer advocacy group tests 845 new cars and finds they have an average of 25.1 miles per gallon.

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.

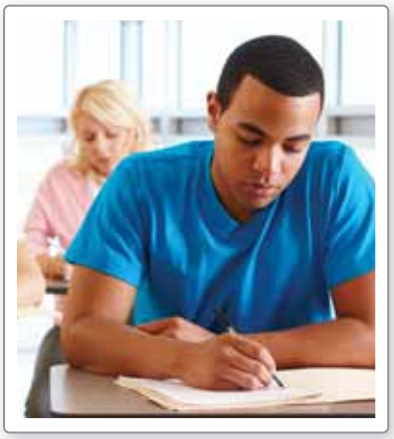


b. The population consists of the gasoline mileages of all new cars sold in the United States, and the sample consists of the gasoline mileages of the 845 new cars tested by the group. Notice in the diagram that the sample is a subset of the gasoline mileages of all new cars in the United States. The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

UNDERSTANDING MATHEMATICAL TERMS

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

Analyzing Hypotheses

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

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

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

EXAMPLE 3 Analyzing a Hypothesis

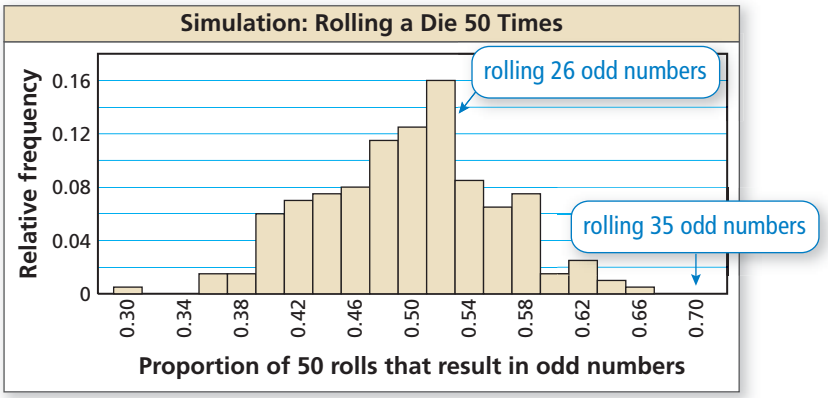
You roll a six-sided die 5 times and do not get an even number. The probability of this happening is $\left(\frac{1}{2}\right)^5 = 0.03125$, so you suspect this die favors odd numbers. The die maker claims the die does not favor odd numbers or even numbers. What should you conclude when you roll the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

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.

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.

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.

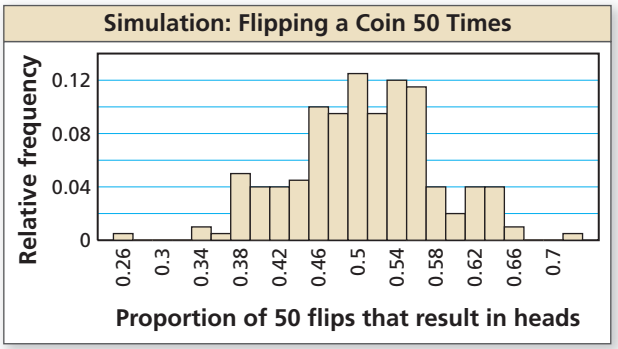
17. **ERROR ANALYSIS** A survey of 1270 high school students found that 965 students felt added stress because of their workload. Describe and correct the error in identifying the population and the sample.

> ✗ 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.

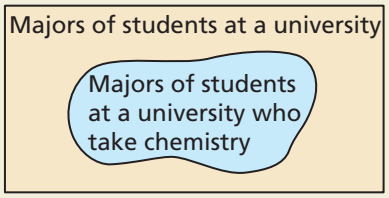
19. MODELING WITH MATHEMATICS You flip a coin 4 times and do not get a tails. You suspect this coin favors heads. The coin maker claims that the coin does not favor heads or tails. You simulate flipping the coin 50 times by repeatedly drawing 200 random samples of size 50. The histogram shows the results. 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.

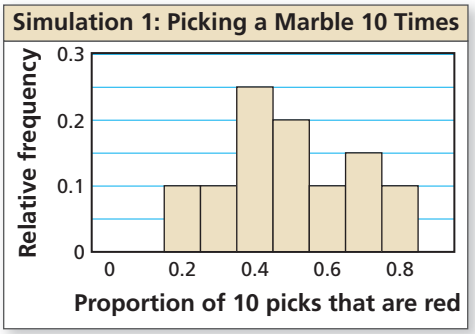
22. HOW DO YOU SEE IT? Use the Venn diagram to identify the population and sample. Explain your reasoning.

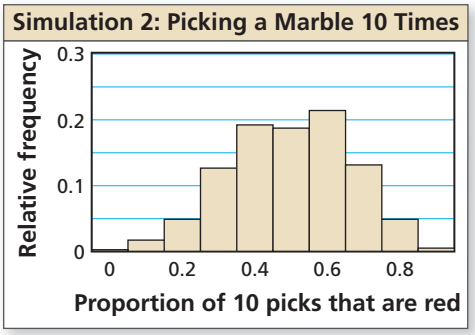


23. OPEN-ENDED Find a newspaper or magazine article that describes a survey. Identify the population and sample. Describe the sample.

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.** $5w - w^2 = -11$