Samples, Populations

11.2

What You Will Learn

- Distinguish between populations and samples.
- Analyze hypotheses

A population is the collection of all data, such as response, 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 this later).

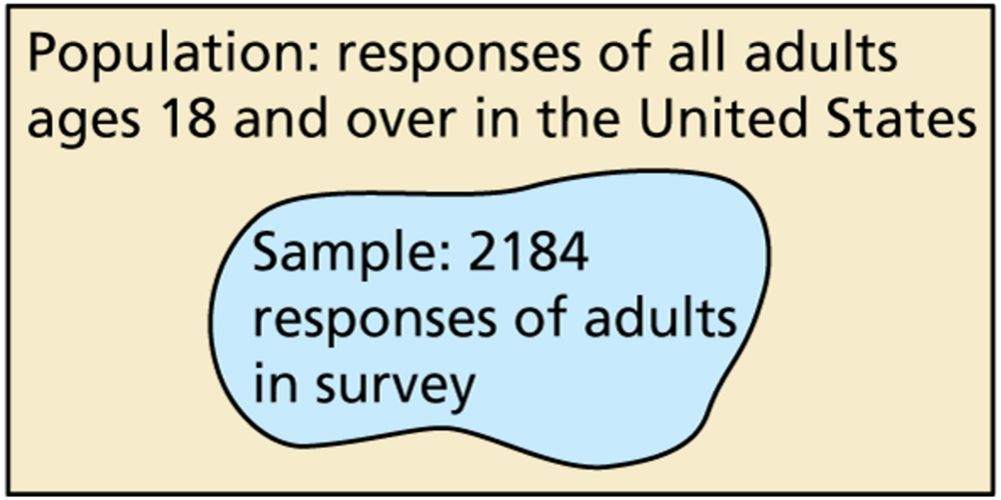
It is important for a sample to be representative of a population so that the sample data can be used to draw conclusions about the population. When the sample is not representative of the population, the conclusions may not be valid. Drawing conclusions about populations is an important use of *statistics.*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

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.

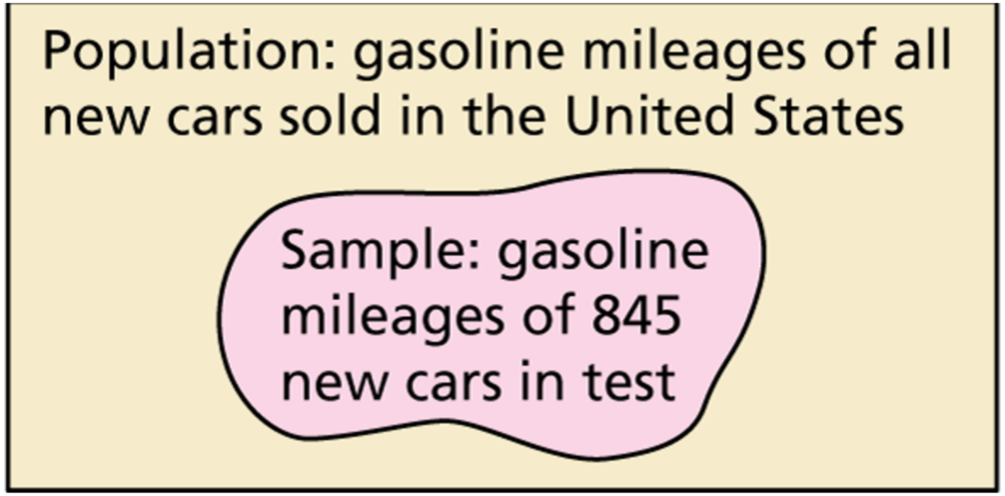
SOLUTION

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. 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.

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.

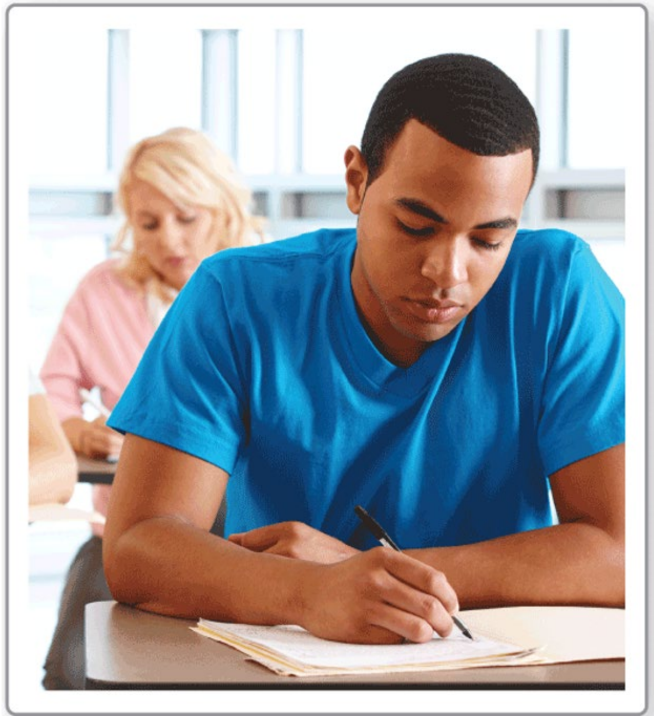


Vocabulary

Parameter: A numerical description of a population

Statistic: A numerical description of a sample

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 can distinguish between a parameter and a statistic



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.

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. 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.

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.

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.

Population: All retail gasoline outlets in the United States
Sample: the 800 retail gasoline outlets the EIA called.

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.

Population: All shoppers in the United States
Sample: The 4464 shoppers in the United States surveyed.

Analyzing Hypotheses

In statistics, a **hypothesis** is a claim about a characteristic of a population.

1. A drug company claims that patients using its weight-loss drug lose an average of 24 pounds in the first three 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%

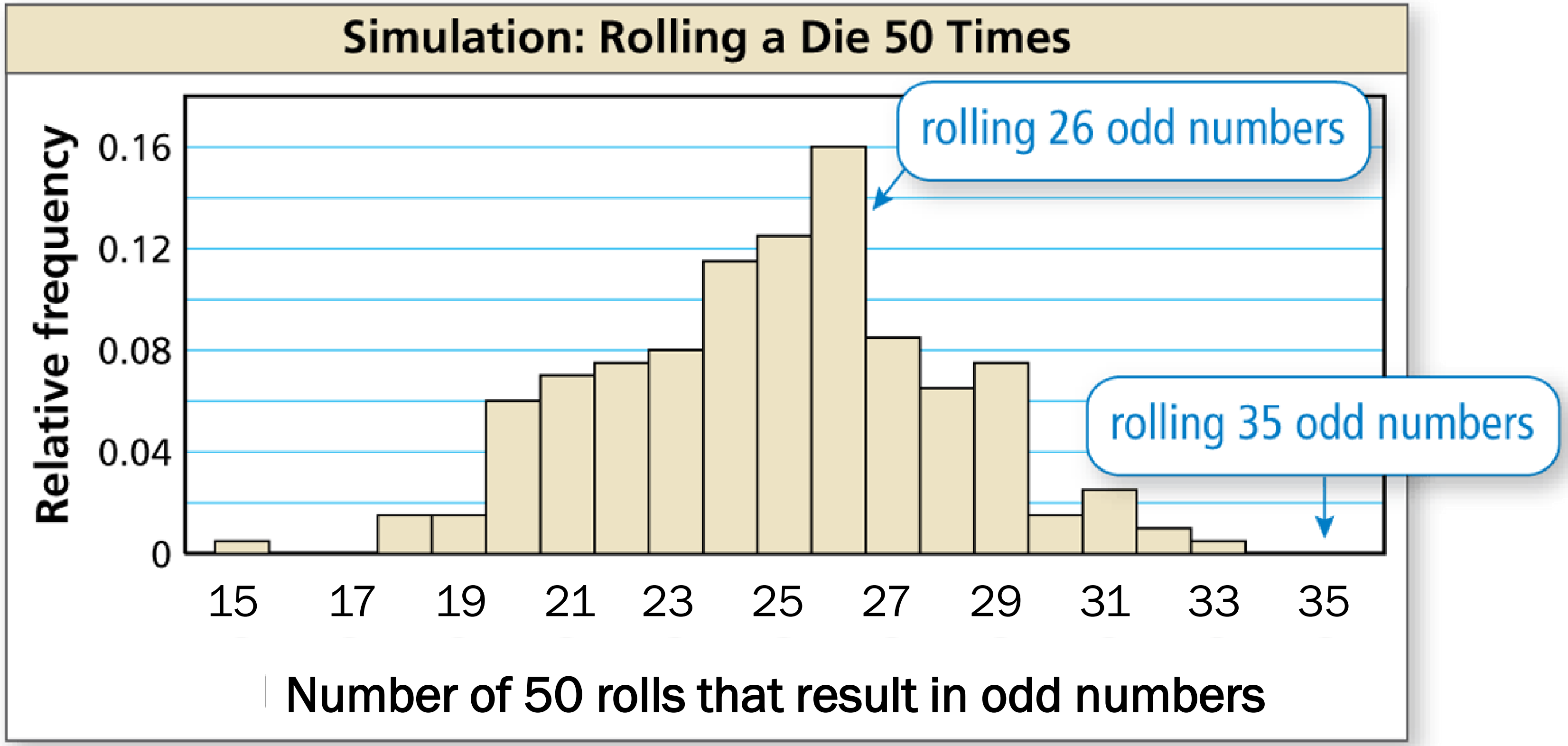
To analyze a hypothesis, you need to distinguish between results that can easily occur by chance and results that are highly unlikely to occur by chance. One way to analyze a hypothesis is to perform a simulation. When the results are highly unlikely to occur, the hypothesis is probably false.

Analyzing a Hypothesis

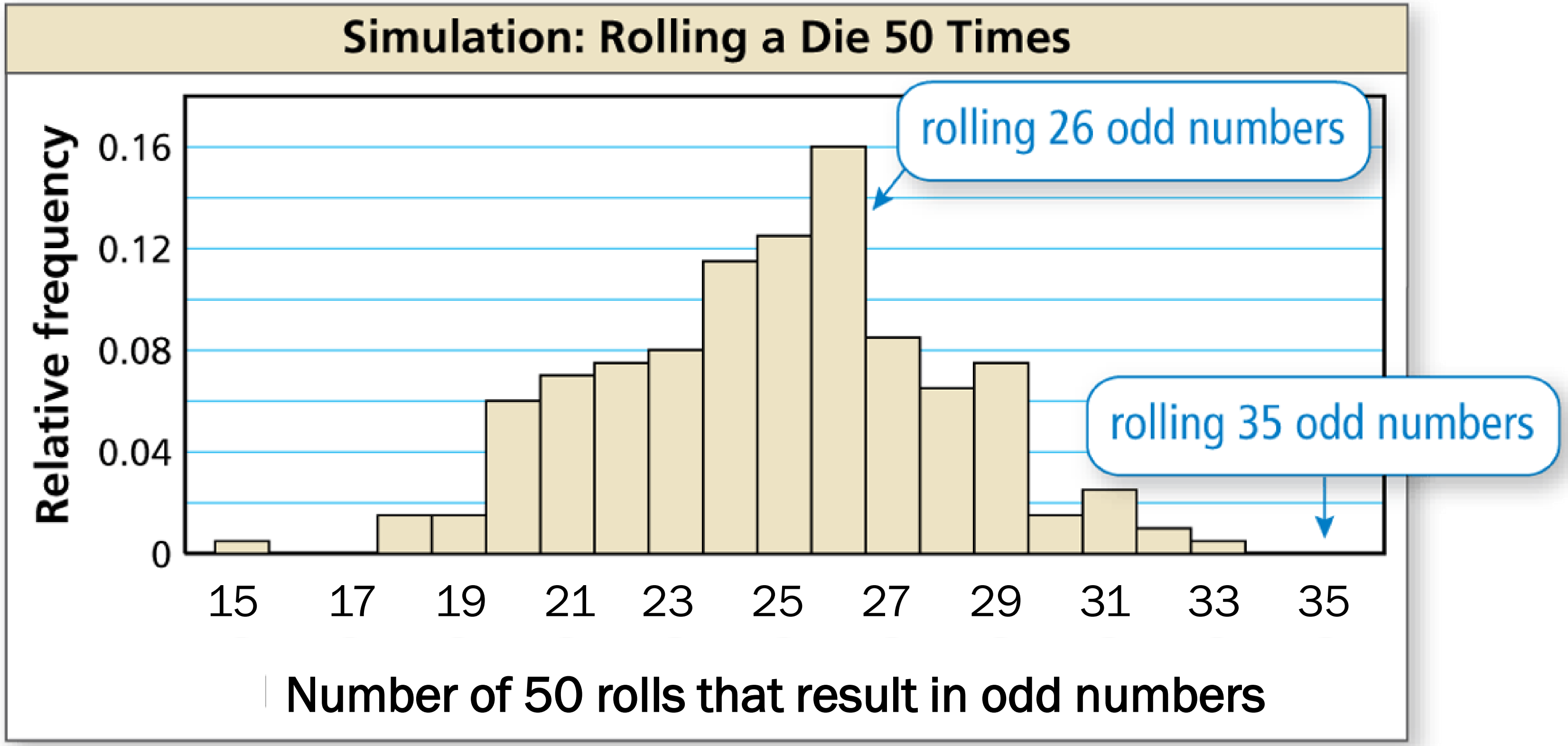
You roll a six-sided die 5 times and do not get an even number. The probability of this happening is $\left(\frac{1}{2}\right)^5 = 0.03125$, so you suspect this die favors odd numbers. The die maker claims the die does not favor odd numbers or even numbers. What should you conclude when you roll the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

SOLUTION

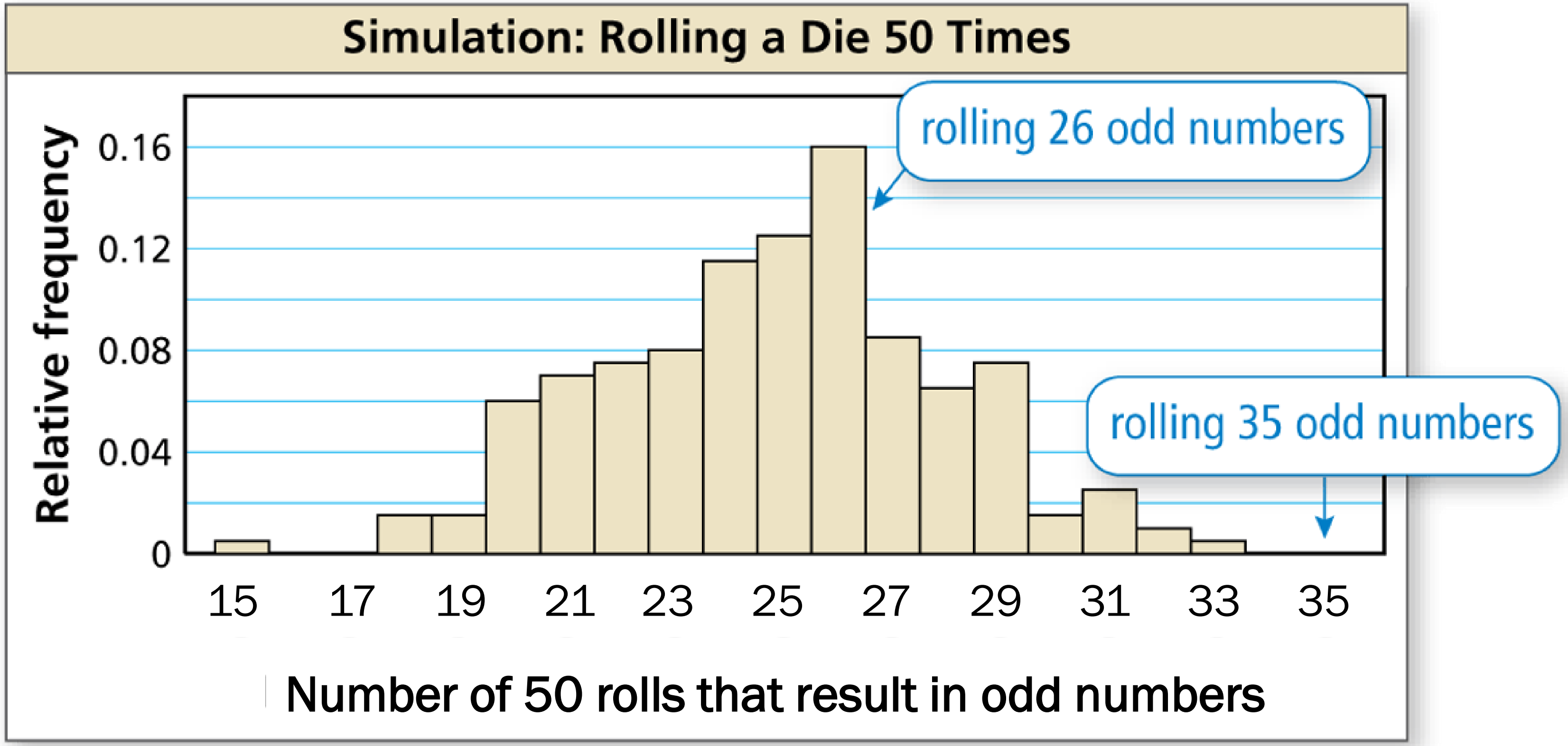
The maker's claim, or hypothesis, is "the die does not favor odd numbers or even numbers." This is the same as saying that the proportion of odd numbers rolled, in the long run, is 0.50. So, assume the probability of rolling an odd number is 0.50. Simulate the rolling of the die by repeatedly drawing 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.



SOLUTION



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. In this simulation, getting 35 odd numbers in 50 rolls apparently did not happen, or happened so few times it does not show up on the plot. Because getting 35 odd numbers is highly unlikely to occur by chance, this simulation confirms that you can conclude the maker's claim is most likely false.

What would you conclude when you roll the actual die 50 times and get (a) 24 odd numbers and (b) 31 odd numbers?

a. The maker's claim is most likely true.

b. The maker's claim is most likely false.

Conclusions:

		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

Yes, I *AM* Captain Obvious!

Homework: Textbook page 607: 5 – 15 odd, 19, 21, 25