

Frequency Distributions

GOAL Identify types of frequency distributions; calculate probabilities and z-scores using normal distributions.

A **random variable** is a variable whose values are determined by the outcome of a probability experiment. The number of heads in 10 tosses of a coin, the height of a randomly selected woman, and the number of members in a randomly selected family are all examples of random variables. In this lesson, you will study two types of random variables: *discrete random variables* and *continuous random variables*.

A **discrete random variable** is a random variable that can take on a countable number of values. The number of sixes when rolling number cubes and the number of heads when tossing coins are examples of discrete random variables. A **continuous random variable** is a random variable that can take on any value in an interval. The amount of rain in a year and the amount of gasoline used by a random commuter driving to and from work are examples of continuous random variables.

EXAMPLE 1 Identifying Discrete and Continuous Random Variables

Identify the given random variable as discrete or continuous.

- The shoe size of a randomly selected person at a bowling alley
- The speed of a randomly selected airplane
- The number of textbooks in a randomly selected student's locker
- The time it takes for a randomly selected student to get to school in the morning

SOLUTION

- Discrete. Shoe size has a countable number of possibilities.
- Continuous. The speed of an airplane can take on any value in the range from 0 to the airplane's top speed.
- Discrete. The number of textbooks has a countable number of possibilities.
- Continuous. Time can take on any value in a certain interval.

Activity

Investigating Relative Frequencies

- Roll two number cubes 50 times. Record the sums of the resulting numbers and the corresponding frequencies in a frequency table.
- Find the *relative frequency* of each sum by dividing the frequency of each sum by the total number of rolls. Record the results in your table.
- Make a histogram with the relative frequencies as the heights of the bars. This type of graph is called a *relative frequency histogram*.
- Compare your histogram to your classmates' histograms. How are the histograms similar? How are they different?

In the activity, you found the distribution of the relative frequencies of the sums that resulted from rolling two number cubes 50 times. You can find the *probability distribution* using the theoretical probability of rolling two number cubes.

A **probability distribution** shows the probabilities of the possible values of a random variable. Probability distributions of discrete random variables are often shown as histograms. The sum of the probabilities is 1.

EXAMPLE 2 Making a Probability Distribution Histogram

Two number cubes are rolled and the sum is recorded.

- Find the theoretical probability of each possible sum.
- Make a histogram of the probability distribution.

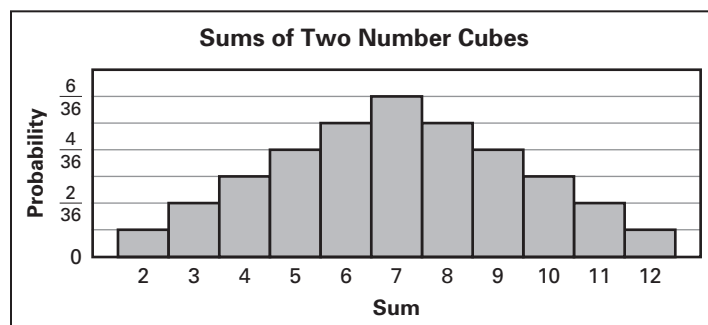
SOLUTION

- Make a table of all of the possible sums and their corresponding probabilities. For example, of 36 possible rolls, there is only 1 roll that results in a sum of 2.

Therefore, the probability of rolling a sum of 2 is $\frac{1}{36}$.

Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- Make a histogram with the probabilities as the heights of the bars.



CHECK Examples 1 and 2

Identify the given random variable as *discrete* or *continuous*.

- The temperature in your city on a randomly selected day
- The number of goals a randomly selected soccer player has scored

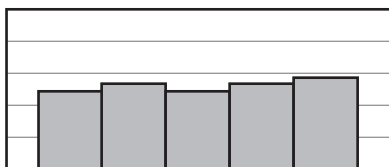
In Exercises 3 and 4, use the following information.

A coin is tossed 4 times and the number of heads is recorded.

- Find the theoretical probability of getting 0, 1, 2, 3, and 4 heads in the 4 tosses of a coin.
- Make a histogram of the probability distribution.

A **frequency distribution** shows the frequencies of the outcomes of an event. Frequency distributions are categorized by the general shape of their histograms.

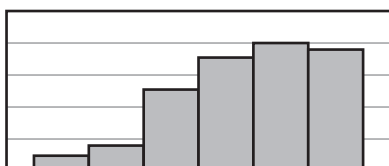
Types of Distributions



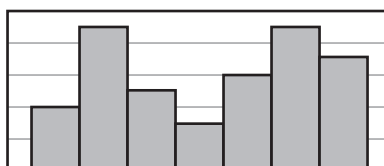
In a **uniform distribution**, all of the intervals have approximately equal frequencies.



A **mound-shaped distribution** is approximately symmetrical. The frequencies of the intervals generally increase from each end toward the distribution's center.

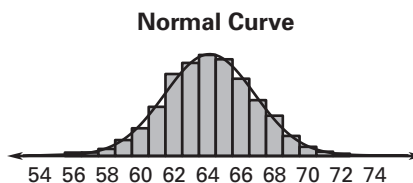
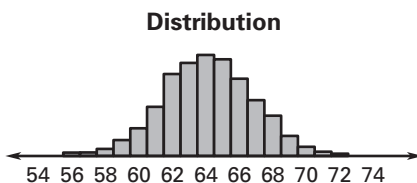


In a **skewed distribution**, the interval or group of intervals that contains the greatest frequency or frequencies is near one end.

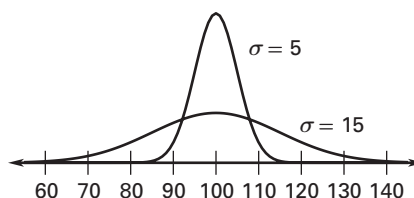


A **bimodal distribution** has two distinct intervals or groups of intervals that contain the greatest frequencies.

Below, the histogram on the left shows the mound-shaped distribution of the heights (in inches) of women in the United States. When a continuous random variable like women's heights has a mound-shaped distribution, you can approximate it with a special curve called a **normal curve**. (Normal curves are sometimes also used to approximate mound-shaped distributions of discrete random variables.) A normal curve is symmetrical, bell-shaped, and has a single peak at its center, which corresponds to the mean. The graph on the right shows a normal curve drawn over the histogram.



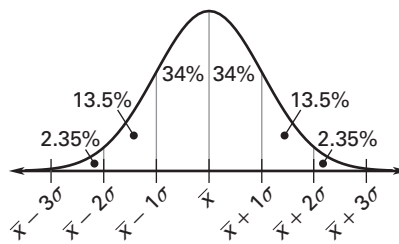
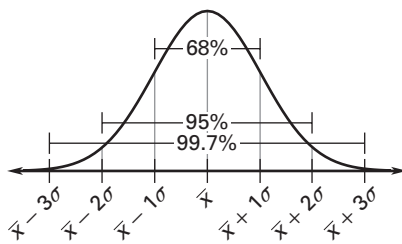
Normal curves can be distinguished by the mean and standard deviation that correspond to the curve. The diagram at the right shows two normal curves. The data that are represented by the curves have the same mean, but different standard deviations. Notice how the standard deviation affects the shape of the curve.



Distributions that can be represented with histograms that follow normal curves are **normal distributions**. All normal distributions follow the *68-95-99.7 rule*. This rule is also known as the *empirical rule*.

68-95-99.7 Rule for Normal Distributions

- About 68% of the data are within 1 standard deviation of the mean.
- About 95% of the data are within 2 standard deviations of the mean.
- About 99.7% of the data are within 3 standard deviations of the mean.



EXAMPLE 3 Using the 68-95-99.7 Rule

The heights of adult women in the United States are normally distributed with a mean of 64 inches and a standard deviation of 2.7 inches.

- Sketch a normal curve to show the distribution of heights. Label the mean and the heights that are 1, 2, and 3 standard deviations from the mean.
- What percent of women in the United States are 61.3 to 66.7 inches tall?
- What percent of women in the United States are 64 to 69.4 inches tall?

SOLUTION

- Sketch a normal curve. Label 64 inches as the mean. Then find and label the heights that are 1, 2, and 3 standard deviations from the mean.

$$\bar{x} - 1\sigma = 64 - 1(2.7) = 61.3 \text{ in.} \quad \bar{x} + 1\sigma = 64 + 1(2.7) = 66.7 \text{ in.}$$

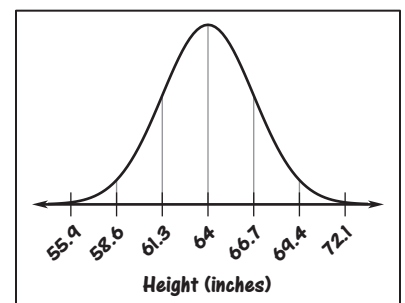
$$\bar{x} - 2\sigma = 64 - 2(2.7) = 58.6 \text{ in.} \quad \bar{x} + 2\sigma = 64 + 2(2.7) = 69.4 \text{ in.}$$

$$\bar{x} - 3\sigma = 64 - 3(2.7) = 55.9 \text{ in.} \quad \bar{x} + 3\sigma = 64 + 3(2.7) = 72.1 \text{ in.}$$

- Both 61.3 inches and 66.7 inches are 1 standard deviation from the mean. Using the 68-95-99.7 rule, you know that about 68% of the data is within 1 standard deviation, so about 68% of women have a height between 61.3 and 66.7 inches.
- About 34% of the women are between 64 inches ($\bar{x}$) and 66.7 inches ($\bar{x} + 1\sigma$), and about 13.5% of the women are between 66.7 ($\bar{x} + 1\sigma$) and 69.4 inches ($\bar{x} + 2\sigma$). To find the percent of women between 64 and 69.4 inches, add these percents.

$$34\% + 13.5\% = 47.5\%$$

So, about 47.5% of women in the United States are 64 to 69.4 inches tall.



A **z-score** indicates how many standard deviations above or below the mean a particular value falls. It can be used to compare values from two different distributions. You can calculate z-scores using the following formula.

$$z\text{-score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \quad \text{or} \quad z = \frac{x - \bar{x}}{\sigma}$$

EXAMPLE 4 Using z-scores

The heights of adult men and women in the United States are normally distributed.

Men: mean = 69 inches, standard deviation = 2.8 inches

Women: mean = 64 inches, standard deviation = 2.7 inches

Justin is 67 inches tall and Ann is 65 inches tall. Find and compare the z-scores for Justin's and Ann's heights.

SOLUTION

$$\text{Justin: } z = \frac{x - \bar{x}}{\sigma} = \frac{67 - 69}{2.8} \approx -0.71 \quad \text{Ann: } z = \frac{x - \bar{x}}{\sigma} = \frac{65 - 64}{2.7} \approx 0.37$$

Justin's height is 0.71 standard deviation below the mean and Ann's height is 0.37 standard deviation above the mean. Ann's height is closer to the mean.



CHECK Examples 3 and 4

In Exercises 5–8, use the following information.

The scores on a college entrance test are normally distributed with a mean of 19 and a standard deviation of 4.5.

5. Sketch a normal curve to show the distribution of scores. Label the mean and the scores that are 1, 2, and 3 standard deviations from the mean.
6. Between what two values does the middle 68% of the scores on the test fall?
7. What percent of scores will be between 10 and 28?
8. Maria's score on the test was 25. Find the z-score for Maria's test score.

EXERCISES

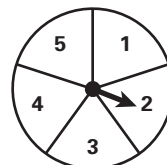
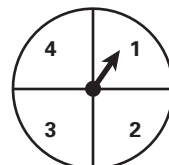
Identify the given random variable as *discrete* or *continuous*.

1. The number of people attending a randomly selected concert
2. The amount of milk produced by a randomly selected cow
3. The number of cars in a parking lot on a randomly selected day
4. The weight of a randomly selected dog

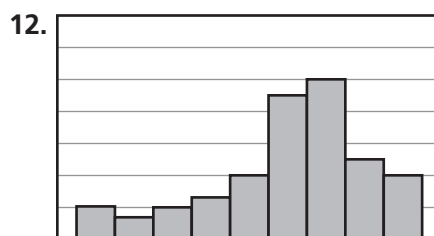
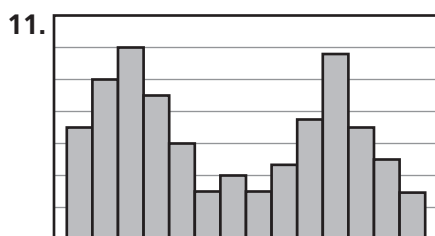
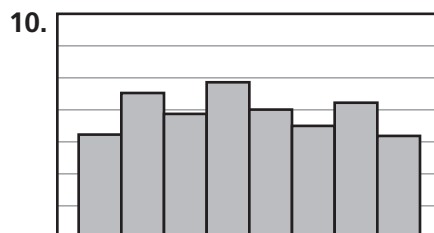
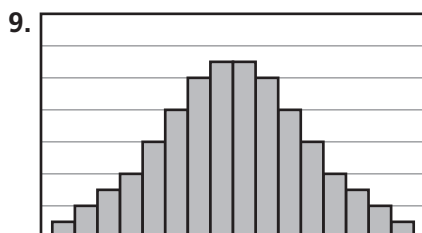
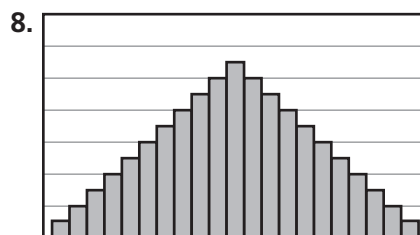
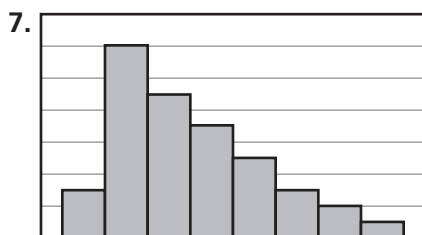
In Exercises 5 and 6, use the following information.

Suppose you spin both of the spinners at the right and record the sum of the numbers. The spinners are each divided into equal parts.

5. Find the theoretical probability of each possible sum.
6. Make a histogram of the probability distribution.



Tell whether the distribution is *uniform*, *mound-shaped*, *skewed*, or *bimodal*.



In Exercises 13–15, use the following information.

The table below shows the distribution of the number of people in a household in the United States in 2000.

Number of people	1	2	3	4	5	6	7 or more
Percent	25.5%	33.1%	16.4%	14.6%	6.7%	2.3%	?

13. Find the percent of households that have 7 people or more.

14. Make a relative frequency histogram.

15. Describe the shape of the distribution.

In Exercises 16–18, use the following information.

Forty-two students were asked how many hours they slept the night before the first day of school. The results are shown in the table below.

Number of hours	5	6	7	8	9	10	11	12
Frequency	1	0	3	8	10	8	8	4

16. Find the relative frequency for each value in the table.

17. Make a relative frequency histogram.

18. Describe the shape of the distribution.

In Exercises 19–22, use the following information.

A company produces bottles of juice that should contain 12 fluid ounces. The actual numbers of fluid ounces are normally distributed with a mean of 12 fluid ounces and a standard deviation of 0.09 fluid ounce.

19. Sketch a normal curve to show the distribution of the numbers of fluid ounces. Label the mean and the numbers of fluid ounces 1, 2, and 3 standard deviations from the mean.
20. Between what two capacities will the middle 95% of the bottles fall?
21. What percent of the bottles will contain more than 12 fluid ounces?
22. A particular bottle contains 11.7 fluid ounces. What is the z -score for this bottle?

In Exercises 23 and 24, use the following information.

The mean daytime high temperatures during January in Rio de Janeiro, Brazil, are normally distributed with a mean of 79.4°F and a standard deviation of 1.6°F .

23. During a typical January, what percent of days would you expect the high temperature to be between 77.8°F and 81°F ?
24. During a typical January, what percent of days would you expect the high temperature to be less than 81°F ?

In Exercises 25–28, use the following information.

The weights of granola bars are normally distributed with a mean of 2.2 ounces and a standard deviation of 0.05 ounce.

25. What percent of granola bars weigh between 2.05 and 2.35 ounces?
26. What percent of the granola bars weigh between 2.2 and 2.25 ounces?
27. What percent of the granola bars weigh less than 2.15 ounces?
28. Any granola bar that weighs less than 2.05 ounces or more than 2.35 ounces can be considered an outlier. What percent of the granola bars can be considered outliers?

In Exercises 29–33, use the following information.

Laura and Max took college entrance tests. The scores on the test that Laura took are normally distributed with a mean of 20 points and a standard deviation of 4.2 points. The scores on the test that Max took are normally distributed with a mean of 500 points and a standard deviation of 90 points.

29. Laura's score was 28. How many points above the mean did Laura score?
30. Max's score was 600. How many points above the mean did Max score?
31. Find the z -score for Laura's entrance test score.
32. Find the z -score for Max's entrance test score.
33. Which student scored better on his or her college entrance test? Explain.
34. The light bulbs that a company produces have life spans that are normally distributed with a mean of 2000 hours and standard deviation of 190 hours.
 - a. What do you think is the median life span of a light bulb? Explain your answer in terms of the shape of the normal curve.
 - b. Suppose you measure the life spans of 10,000 light bulbs and round the life spans to the nearest hour. What would you expect to be the mode of the data? Explain your answer in terms of the shape of the normal curve.

In Exercises 35–37, use the following information.

Given data that are normally distributed, a value is generally considered to be an outlier if it is more than 3 standard deviations from the mean.

- 35.** Based on the 68-95-99.7 rule, what is the probability that a randomly-chosen value from a normally-distributed data set will be an outlier?
- 36.** What can you say about the z -score for a value that is an outlier in a normally-distributed data set?
- 37.** The length of bolts made at a factory are normally distributed with a mean of 3.45 cm and a standard deviation of 0.04 cm. A quality-control inspector finds bolts with lengths of 3.56 cm and 3.32 cm. Find the z -score for each bolt. Is either bolt an outlier? Explain.

In Exercises 38–40, use the following information.

A *skewed right* distribution is one in which the “tail” is on the right side. A *skewed left* distribution is one in which the tail is on the left side. Consider the distributions at the right, which show the ages of teenagers who attended three workshops at a local school.

- 38.** Describe each distribution as *skewed right*, *skewed left*, or *mound-shaped*.

- 39.** The data sets for the distributions are given below. Find the mean, median, mode, and range for each distribution.

Passing a driving test: 13, 14, 14, 14, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 18, 18, 18, 19

Adjusting to a new school: 13, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 16, 16, 16, 17, 17, 18, 19

Choosing a college: 13, 14, 15, 15, 16, 16, 16, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18, 18, 19, 19, 19, 19, 19

- 40.** Describe the relative positions of the mean, median, and mode for each distribution. That is, for each distribution, explain which of these measures lies to the left or right of the other measures.

