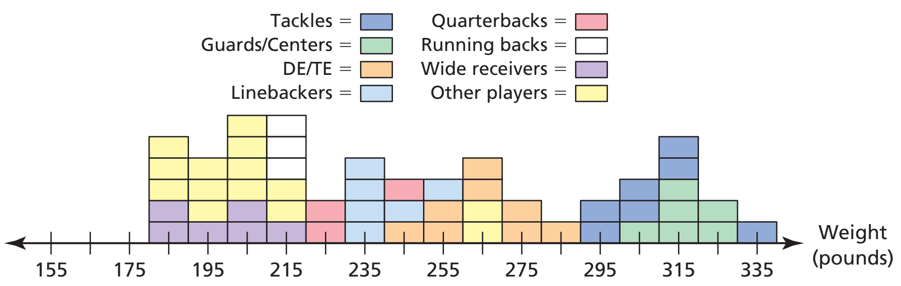
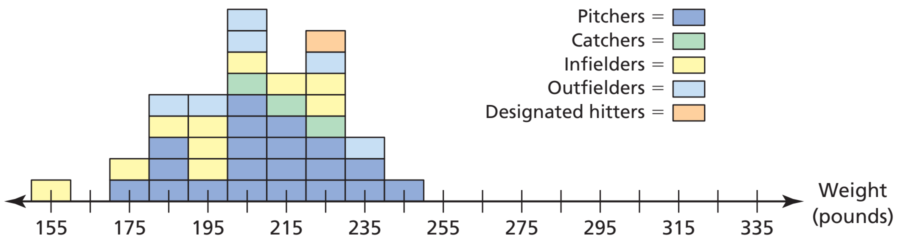
11.1 Measures of Center and Variation

Essential Question How can you describe the variation of a data set?

EXPLORATION 1 Describing the Variation of Data

Work with a partner. The graphs show the weights of the players on a professional football team and a professional baseball team.





CONSTRUCTING VIABLE ARGUMENTS

To be proficient in math, you need to reason inductively about data, making plausible arguments that take into account the context from which the data arose.

a. Describe the data in each graph in terms of how much the weights vary from the mean. Explain your reasoning.

b. Compare how much the weights of the players on the football team vary from the mean to how much the weights of the players on the baseball team vary from the mean.

c. Does there appear to be a correlation between the body weights and the positions of players in professional football? in professional baseball? Explain.

EXPLORATION 2 Describing the Variation of Data

Work with a partner. The weights (in pounds) of the players on a professional basketball team by position are as follows.

Power forwards: 235, 255, 295, 245; small forwards: 235, 235;
centers: 255, 245, 325; point guards: 205, 185, 205; shooting guards: 205, 215, 185

Make a graph that represents the weights and positions of the players. Does there appear to be a correlation between the body weights and the positions of players in professional basketball? Explain your reasoning.

Communicate Your Answer

3. How can you describe the variation of a data set?

11.1 Lesson

Core Vocabulary

- measure of center, *p. 586*
- mean, *p. 586*
- median, *p. 586*
- mode, *p. 586*
- outlier, *p. 587*
- measure of variation, *p. 587*
- range, *p. 587*
- standard deviation, *p. 588*
- data transformation, *p. 589*

What You Will Learn

- Compare the mean, median, and mode of a data set.
- Find the range and standard deviation of a data set.
- Identify the effects of transformations on data.

Comparing the Mean, Median, and Mode

A **measure of center** is a measure that represents the center, or typical value, of a data set. The *mean*, *median*, and *mode* are measures of center.

Core Concept

Mean

The **mean** of a numerical data set is the sum of the data divided by the number of data values. The symbol $\overline{x}$ represents the mean. It is read as "*x*-bar."

Median

The **median** of a numerical data set is the middle number when the values are written in numerical order. When a data set has an even number of values, the median is the mean of the two middle values.

Mode

The **mode** of a data set is the value or values that occur most often. There may be one mode, no mode, or more than one mode.

EXAMPLE 1 Comparing Measures of Center

An amusement park hires students for the summer. The students' hourly wages are shown in the table.

Students' Hourly Wages	
\$16.50	\$8.25
\$8.75	\$8.45
\$8.65	\$8.25
\$9.10	\$9.25

a. Find the mean, median, and mode of the hourly wages.

b. Which measure of center best represents the data? Explain.

SOLUTION

a. Mean $\overline{x} = \dfrac{16.5 + 8.75 + 8.65 + 9.1 + 8.25 + 8.45 + 8.25 + 9.25}{8} = 9.65$

Median 8.25, 8.25, 8.45, 8.65, 8.75, 9.10, 9.25, 16.50 — Order the data.

$\dfrac{17.4}{2} = 8.7$ — Mean of two middle values

Mode 8.25, 8.25, 8.45, 8.65, 8.75, 9.10, 9.25, 16.50 — 8.25 occurs most often.

▶ The mean is \$9.65, the median is \$8.70, and the mode is \$8.25.

b. 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.

STUDY TIP

Mode is the only measure of center that can represent a nonnumerical data set.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

1. WHAT IF? The park hires another student at an hourly wage of \$8.45.
(a) How does this additional value affect the mean, median, and mode? Explain.
(b) Which measure of center best represents the data? Explain.

An **outlier** is a data value that is much greater than or much less than the other values in a data set.

EXAMPLE 2 Removing an Outlier

Consider the data in Example 1. (a) Identify the outlier. How does the outlier affect the mean, median, and mode? (b) Describe one possible explanation for the outlier.

SOLUTION

a. The value $16.50 is much greater than the other wages. It is the outlier. Find the mean, median, and mode without the outlier.

Mean $\bar{x} = \dfrac{8.75 + 8.65 + 9.1 + 8.25 + 8.45 + 8.25 + 9.25}{7} \approx 8.67$

Median 8.25, 8.25, 8.45, 8.65, 8.75, 9.10, 9.25 — The middle value is 8.65.

Mode 8.25, 8.25, 8.45, 8.65, 8.75, 9.10, 9.25 — The mode is 8.25.

▶ When you remove the outlier, the mean decreases $9.65 − $8.67 = $0.98, the median decreases $8.70 − $8.65 = $0.05, and the mode is the same.

b. The outlier could be a student who is hired to maintain the park's website, while the other students could be game attendants.

STUDY TIP

Outliers usually have the greatest effect on the mean.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

Annual Salaries	
$32,000	$42,000
$41,000	$38,000
$38,000	$45,000
$72,000	$35,000

2. The table shows the annual salaries of the employees of an auto repair service. (a) Identify the outlier. How does the outlier affect the mean, median, and mode? (b) Describe one possible explanation for the outlier.

Finding the Range and Standard Deviation

A **measure of variation** is a measure that describes the spread, or distribution, of a data set. One measure of variation is the *range*. The **range** of a data set is the difference of the greatest value and the least value.

EXAMPLE 3 Finding a Range

Two reality cooking shows select 12 contestants each. The ages of the contestants are shown in the tables. Find the range of the ages for each show. Compare your results.

Show A	
Ages	
20	29
19	22
25	27
27	29
30	20
21	31

Show B	
Ages	
25	19
20	27
22	25
27	22
48	21
32	24

SOLUTION

Show A 19, 20, 20, 21, 22, 25, 27, 27, 29, 29, 30, 31 — Order the data.

So, the range is 31 − 19, or 12 years.

Show B 19, 20, 21, 22, 22, 24, 25, 25, 27, 27, 32, 48 — Order the data.

So, the range is 48 − 19, or 29 years.

▶ The range of the ages for Show A is 12 years, and the range of the ages for Show B is 29 years. So, the ages for Show B are more spread out.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

3. After the first week, the 25-year-old is voted off Show A and the 48-year-old is voted off Show B. How does this affect the range of the ages of the remaining contestants on each show in Example 3? Explain.

A disadvantage of using the range to describe the spread of a data set is that it uses only two data values. A measure of variation that uses all the values of a data set is the *standard deviation*.

Core Concept

Standard Deviation

The **standard deviation** of a numerical data set is a measure of how much a typical value in the data set differs from the mean. The symbol σ represents the standard deviation. It is read as "sigma." It is given by

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

where n is the number of values in the data set. The deviation of a data value x is the difference of the data value and the mean of the data set, $x - \overline{x}$.

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

Step 2 Find the deviation of each data value, $x - \overline{x}$.

Step 3 Square each deviation, $(x - \overline{x})^2$.

Step 4 Find the mean of the squared deviations. This is called the *variance*.

Step 5 Take the square root of the variance.

REMEMBER

An ellipsis " $\cdots$ " indicates that a pattern continues.

A small standard deviation means that the data are clustered around the mean. A large standard deviation means that the data are more spread out.

EXAMPLE 4 Finding a Standard Deviation

Find the standard deviation of the ages for Show A in Example 3. Use a table to organize your work. Interpret your result.

SOLUTION

x	$\overline{x}$	$x - \overline{x}$	$(x - \overline{x})^2$
20	25	−5	25
29	25	4	16
19	25	−6	36
22	25	−3	9
25	25	0	0
27	25	2	4
27	25	2	4
29	25	4	16
30	25	5	25
20	25	−5	25
21	25	−4	16
31	25	6	36

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

$$\overline{x} = \frac{300}{12} = 25$$

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

Step 3 Square each deviation, $(x - \overline{x})^2$, as shown.

Step 4 Find the mean of the squared deviations, or variance.

$$\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n} = \frac{25 + 16 + \cdots + 36}{12} = \frac{212}{12} \approx 17.7$$

Step 5 Use a calculator to take the square root of the variance.

$$\sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n}} = \sqrt{\frac{212}{12}} \approx 4.2$$

▶ The standard deviation is about 4.2. This means that the typical age of a contestant on Show A differs from the mean by about 4.2 years.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

4. Find the standard deviation of the ages for Show B in Example 3. Interpret your result.

5. Compare the standard deviations for Show A and Show B. What can you conclude?

Effects of Data Transformations

A **data transformation** is a procedure that uses a mathematical operation to change a data set into a different data set.

STUDY TIP

The standard deviation stays the same because the amount by which each data value deviates from the mean stays the same.

Core Concept

Data Transformations Using Addition

When a real number k is added to each value in a numerical data set

- the measures of center of the new data set can be found by adding k to the original measures of center.
- the measures of variation of the new data set are the *same* as the original measures of variation.

Data Transformations Using Multiplication

When each value in a numerical data set is multiplied by a real number k, where $k > 0$, the measures of center and variation can be found by multiplying the original measures by k.

EXAMPLE 5 Real-Life Application

Consider the data in Example 1. (a) Find the mean, median, mode, range, and standard deviation when each hourly wage increases by \$0.50. (b) Find the mean, median, mode, range, and standard deviation when each hourly wage increases by 10%.

SOLUTION

Students' Hourly Wages	
\$17.00	\$8.75
\$9.25	\$8.95
\$9.15	\$8.75
\$9.60	\$9.75

a. Method 1 Make a new table by adding \$0.50 to each hourly wage. Find the mean, median, mode, range, and standard deviation of the new data set.

▶ Mean: \$10.15 Median: \$9.20 Mode: \$8.75
Range: \$8.25 Standard deviation: \$2.61

Method 2 Find the mean, median, mode, range, and standard deviation of the original data set.

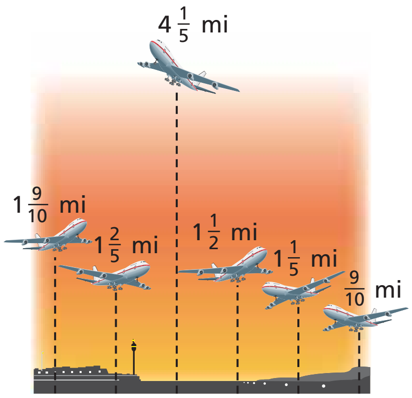
Mean: \$9.65 Median: \$8.70 Mode: \$8.25 From Example 1
Range: \$8.25 Standard deviation: \$2.61

Add \$0.50 to the mean, median, and mode. The range and standard deviation are the same as the original range and standard deviation.

▶ Mean: \$10.15 Median: \$9.20 Mode: \$8.75
Range: \$8.25 Standard deviation: \$2.61

b. Increasing by 10% means to multiply by 1.1. So, multiply the original mean, median, mode, range, and standard deviation from Method 2 of part (a) by 1.1.

▶ Mean: \$10.62 Median: \$9.57 Mode: \$9.08
Range: \$9.08 Standard deviation: \$2.87



Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

6. Find the mean, median, mode, range, and standard deviation of the altitudes of the airplanes when each altitude increases by $1\frac{1}{2}$ miles.

11.1 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **VOCABULARY** In a data set, what does a measure of center represent? What does a measure of variation describe?

2. **WRITING** Describe how removing an outlier from a data set affects the mean of the data set.

3. **OPEN-ENDED** Create a data set that has more than one mode.

4. **REASONING** What is an advantage of using the range to describe a data set? Why do you think the standard deviation is considered a more reliable measure of variation than the range?

Monitoring Progress and Modeling with Mathematics

In Exercises 5–8, (a) find the mean, median, and mode of the data set and (b) determine which measure of center best represents the data. Explain. *(See Example 1.)*

5. 3, 5, 1, 5, 1, 1, 2, 3, 15

6. 12, 9, 17, 15, 10

7. 13, 30, 16, 19, 20, 22, 25, 31

8. 14, 15, 3, 15, 14, 14, 18, 15, 8, 16

9. **ANALYZING DATA** The table shows the lengths of nine movies.

Movie Lengths (hours)		
$1\frac{1}{3}$	$1\frac{2}{3}$	2
3	$2\frac{1}{3}$	$1\frac{2}{3}$
2	2	$1\frac{2}{3}$

   a. Find the mean, median, and mode of the lengths.

   b. Which measure of center best represents the data? Explain.

10. **ANALYZING DATA** The table shows the daily changes in the value of a stock over 12 days.

Changes in Stock Value (dollars)			
1.05	2.03	−13.78	−2.41
2.64	0.67	4.02	1.39
0.66	−0.28	−3.01	2.20

   a. Find the mean, median, and mode of the changes in stock value.

   b. Which measure of center best represents the data? Explain.

   c. On the 13th day, the value of the stock increases by \$4.28. How does this additional value affect the mean, median, and mode? Explain.

In Exercises 11–14, find the value of x.

11. 2, 8, 9, 7, 6, x; The mean is 6.

12. 12.5, −10, −7.5, x; The mean is 11.5.

13. 9, 10, 12, x, 20, 25; The median is 14.

14. 30, 45, x, 100; The median is 51.

15. **ANALYZING DATA** The table shows the masses of eight polar bears. *(See Example 2.)*

Masses (kilograms)			
455	262	471	358
364	553	62	351

   a. Identify the outlier. How does the outlier affect the mean, median, and mode?

   b. Describe one possible explanation for the outlier.

16. **ANALYZING DATA** The sizes of emails (in kilobytes) in your inbox are 2, 3, 5, 2, 1, 46, 3, 7, 2, and 1.

   a. Identify the outlier. How does the outlier affect the mean, median, and mode?

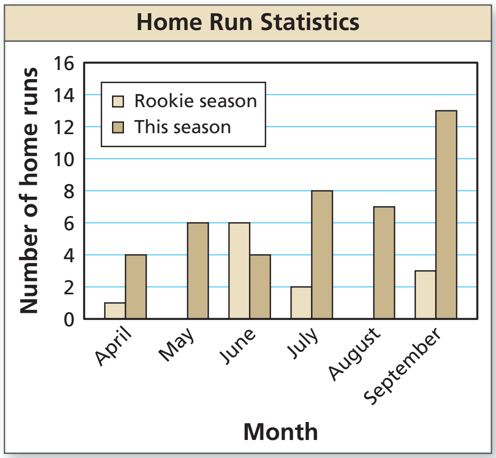
   b. Describe one possible explanation for the outlier.

17. **ANALYZING DATA** The scores of two golfers are shown. Find the range of the scores for each golfer. Compare your results. *(See Example 3.)*

Golfer A	
83	88
84	95
91	89
90	87
98	95

Golfer B	
89	87
93	95
92	94
88	91
89	92

18. **ANALYZING DATA** The graph shows a player's monthly home run totals in two seasons. Find the range of the number of home runs for each season. Compare your results.



In Exercises 19–22, find (a) the range and (b) the standard deviation of the data set.

19. 40, 35, 45, 55, 60

20. 141, 116, 117, 135, 126, 121

21. 0.5, 2.0, 2.5, 1.5, 1.0, 1.5

22. 8.2, 10.1, 2.6, 4.8, 2.4, 5.6, 7.0, 3.3

23. **ANALYZING DATA** Consider the data in Exercise 17. *(See Example 4.)*

a. Find the standard deviation of the scores of Golfer A. Interpret your result.

b. Find the standard deviation of the scores of Golfer B. Interpret your result.

c. Compare the standard deviations for Golfer A and Golfer B. What can you conclude?

24. **ANALYZING DATA** Consider the data in Exercise 18.

a. Find the standard deviation of the monthly home run totals in the player's rookie season. Interpret your result.

b. Find the standard deviation of the monthly home run totals in this season. Interpret your result.

c. Compare the standard deviations for the rookie season and this season. What can you conclude?

In Exercises 25 and 26, find the mean, median, and mode of the data set after the given transformation.

25. In Exercise 5, each data value increases by 4.

26. In Exercise 6, each data value increases by 20%.

27. **TRANSFORMING DATA** Find the values of the measures shown when each value in the data set increases by 14. *(See Example 5.)*

Mean: 62 Median: 55 Mode: 49

Range: 46 Standard deviation: 15.5

28. **TRANSFORMING DATA** Find the values of the measures shown when each value in the data set is multiplied by 0.5.

Mean: 320 Median: 300 Mode: none

Range: 210 Standard deviation: 70.6

29. **ERROR ANALYSIS** Describe and correct the error in finding the median of the data set.

> ✗ 7, 4, 6, 2, 4, 6, 8, 8, 3
>
> The median is 4.

30. **ERROR ANALYSIS** Describe and correct the error in finding the range of the data set after the given transformation.

> ✗ −13, −12, −7, 2, 10, 13
>
> Add 10 to each value.
>
> The range is 26 + 10 = 36.

31. **PROBLEM SOLVING** In a bowling match, the team with the greater mean score wins. The scores of the members of two bowling teams are shown.

Team A: 172, 130, 173, 212
Team B: 136, 184, 168, 192

a. Which team wins the match? If the team with the greater median score wins, is the result the same? Explain.

b. Which team is more consistent? Explain.

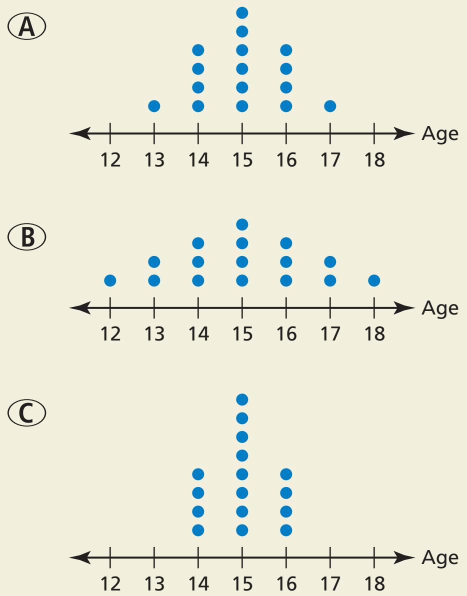
c. In another match between the two teams, all the members of Team A increase their scores by 15 and all the members of Team B increase their scores by 12.5%. Which team wins this match? Explain.

32. **MAKING AN ARGUMENT** Your friend says that when two data sets have the same range, you can assume the data sets have the same standard deviation, because both range and standard deviation are measures of variation. Is your friend correct? Explain.

33. **ANALYZING DATA** The table shows the results of a survey that asked 12 students about their favorite meal. Which measure of center (mean, median, or mode) can be used to describe the data? Explain.

Favorite Meal			
spaghetti	pizza	steak	hamburger
steak	taco	pizza	chili
pizza	chicken	fish	spaghetti

34. **HOW DO YOU SEE IT?** The dot plots show the ages of the members of three different adventure clubs. Without performing calculations, which data set has the greatest standard deviation? Which has the least standard deviation? Explain your reasoning.



35. **REASONING** A data set is described by the measures shown.

Mean: 27　　Median: 32　　Mode: 18

Range: 41　　Standard deviation: 9

Find the mean, median, mode, range, and standard deviation of the data set when each data value is multiplied by 3 and then increased by 8.

36. **CRITICAL THINKING** Can the standard deviation of a data set be 0? Can it be negative? Explain.

37. **USING TOOLS** Measure the heights (in inches) of the students in your class.

a. Find the mean, median, mode, range, and standard deviation of the heights.

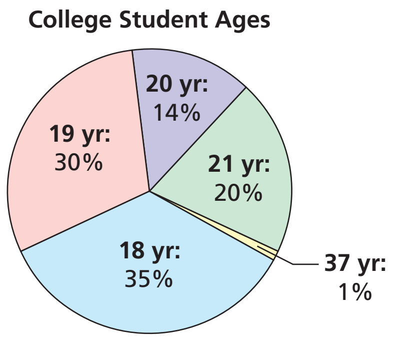
b. A new student who is 7 feet tall joins your class. How would you expect this student's height to affect the measures in part (a)? Verify your answer.

38. **THOUGHT PROVOKING** To find the arithmetic mean of n numbers, divide the sum of the numbers by n. To find the geometric mean of n numbers $a_1, a_2, a_3, \ldots, a_n$, take the nth root of the product of the numbers.

$$\text{geometric mean} = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \cdot \ldots \cdot a_n}$$

Compare the arithmetic mean to the geometric mean of n numbers.

39. **PROBLEM SOLVING** The circle graph shows the distribution of the ages of 200 students in a college Psychology I class.



a. Find the mean, median, and mode of the students' ages.

b. Identify the outliers. How do the outliers affect the mean, median, and mode?

c. Suppose all 200 students take the same Psychology II class exactly 1 year later. Draw a new circle graph that shows the distribution of the ages of this class and find the mean, median, and mode of the students' ages.

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Solve the inequality. *(Section 2.4)*

40. $6x + 1 \leq 4x - 9$　**41.** $-3(3y - 2) < 1 - 9y$　**42.** $2(5c - 4) \geq 5(2c + 8)$　**43.** $4(3 - w) > 3(4w - 4)$

Evaluate the function for the given value of x. *(Section 6.3)*

44. $f(x) = 4^x; x = 3$　**45.** $f(x) = 7^x; x = -2$　**46.** $f(x) = 5(2)^x; x = 6$　**47.** $f(x) = -2(3)^x; x = 4$