9.1   Measures of Center and Spread

Essential Question: How can you describe and compare data sets?

⊘ Explore   Exploring Data

Caleb and Kim have bowled three games. Their scores are shown in the chart below.

Name	Game 1	Game 2	Game 3	Average Score
Caleb	151	153	146	
Kim	122	139	189	

Complete the table.

(A) Find Caleb's average score over the three games and enter it in the table.

$$\frac{151 + 153 + 146}{3} = \frac{\boxed{}}{3} = \boxed{}$$

(B) Find Kim's average score over the three games and enter it in the table.

$$\frac{122 + 139 + 189}{3} = \frac{\boxed{}}{3} = \boxed{}$$

(C) How do their average scores compare?

_____

(D) Caleb's / Kim's scores are more consistent.

(E) Caleb's scores are farther from / closer to the average than Kim's.

They bowl a fourth game, where Caleb scores 150 and Kim scores a 175.
How does this affect their averages?

(F) Caleb's average _____.   (G) Kim's average _____.

(H) Does the Game 4 score affect the consistency of their scores? Explain.

_____

1. **Discussion** Is the average an accurate representation of Caleb's bowling?

_____

_____

2. **Discussion** Is the average an accurate representation of Kim's bowling?

_____

_____

🔧 Explain 1   **Measures of Center: Mean and Median**

Two commonly used measures of center for a set of numerical data are the mean and median. Measures of center represent a central or typical value of a data set. The **mean** is the sum of the values in the set divided by the number of values in the set. The **median** is the middle value in a set when the values are arranged in numerical order.

Example 1   **Find the mean and median of each data set.**

Ⓐ The number of text messages that Isaac received each day for a week is shown.

47, 49, 54, 50, 48, 47, 55

Find the mean. Divide the sum by the numbers of data values.

$\frac{350}{7} = 50$. The mean is 50 text messages a day.

Find the median. Rewrite the values in increasing order.

47, 47, 48, (49), 50, 54, 55. The median is 49 text messages a day.

Ⓑ The amount of money Elise earned in tips per day for 6 days is listed below.

$75, $97, $360, $84, $119, $100

Find the mean to the nearest $0.01. Divide the sum by the number of data values.

$\frac{\boxed{}}{6} = \boxed{}$ The mean is _____.

Find the median. Rewrite the values in increasing order.

75, $\boxed{}$, $\boxed{}$, $\boxed{}$, $\boxed{}$, 360

Find the mean of the middle two values. $\frac{\boxed{} + \boxed{}}{2} = \boxed{}$

The median is _____.

Reflect

3. **Discussion** For the data on tips, which measure of center is more accurate in describing the typical value? Explain.
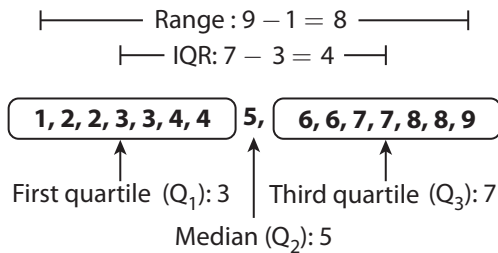
_____

_____

Your Turn

Find the mean and median of each data set.

4. Niles scored 70, 74, 72, 71, 73, and 96 on his 6 geography tests.

5. Raul recorded the following golf scores in his last 7 games.
84, 94, 93, 89, 94, 81, 90

🖉 Explain 2   Measures of Spread: Range and IQR

Measures of spread are used to describe the consistency of data values. They show the distance between data values and their distance from the center of the data. Two commonly used measures of spread for a set of numerical data are the *range* and *interquartile range (IQR)*. The **range** is the difference between the greatest and the least data values. **Quartiles** are values that divide a data set into four equal parts. The **first quartile** (Q_1) is the median of the lower half of the set, the **second quartile** (Q_2) is the median of the whole set, and the **third quartile** (Q_3) is the median of the upper half of the set. The **interquartile range (IQR)** of a data set is the difference between the third and first quartiles. It represents the range of the middle half of the data.

$$\vdash \!\!\!\!—— \text{Range}: 9 - 1 = 8 ——\!\!\!\!\dashv$$
$$\vdash\!\!— \text{IQR}: 7 - 3 = 4 —\!\!\dashv$$

| 1, 2, 2, 3, 3, 4, 4 | 5, | 6, 6, 7, 7, 8, 8, 9 |

First quartile (Q_1): 3   Third quartile (Q_3): 7

Median (Q_2): 5

Example 2   **Find the median, range, and interquartile range for the given data set.**

Ⓐ   The April high temperatures for 5 years in Boston are 77 °F, 86 °F, 84 °F, 93 °F, and 90 °F

Order the data values.

Median: 77, 84, 86, 90, 93   Range: $93 - 77 = 16$
              (median)

Interquartile range:

lower half   upper half
77, 84, 86, 90, 93 $Q_1 = \dfrac{77 + 84}{2} = 80.5$ and $Q_3 = \dfrac{90 + 93}{2} = 91.5$
         (median)

$IQR = Q_3 - Q_1 = 91.5 - 80.5 = 11$

(B) The numbers of runs scored by a softball team in 20 games are given.

3, 4, 8, 12, 7, 5, 4, 12, 3, 9, 11, 4, 14, 8, 2, 10, 3, 10, 9, 7

Order the data values.

2, 3, 3, 3, 4, 4, 4, ☐, ☐, ☐, ☐, ☐, ☐, 9, 10, 10, 11, 12, 12, 14

$\underbrace{}_{\text{median}}$

$\text{Median} = \dfrac{\boxed{} + \boxed{}}{2} = \boxed{}$

$\text{Range} = 14 - 2 = \boxed{}$

Interquartile range.

$Q_1 = \dfrac{\boxed{} + \boxed{}}{2} = 4 \text{ and } Q_3 = \dfrac{\boxed{} + \boxed{}}{2} = \boxed{}$

$IQR = Q_3 - Q_1$

$= \boxed{} - \boxed{}$

$= \boxed{}$

The median is _____.

The range is _____.

The IQR is _____.

6. **Discussion** Why is the IQR less than the range?

_____

_____

_____

_____

Your Turn

Find the median, range, and interquartile range for the given data set.

7. 21, 31, 26, 24, 28, 26

8. The high temperatures in degrees Fahrenheit on 11 days were 68, 71, 75, 74, 75, 71, 73, 71, 72, 74, and 79.

⚙ Explain 3   Measures of Spread: Standard Deviation

Another measure of spread is the **standard deviation**, which represents the average of the distance between individual data values and the mean.

The formula for finding the standard deviation of the data set $\{x_1, x_2, x_2, x_2 \cdots, x_n\}$, with n elements and mean x, is shown below.

$$\text{standard deviation} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

Example 3

(A) Find the standard deviation of 77, 86, 84, 93, 90.

Find the mean.

$$\begin{aligned}\text{mean} &= \frac{77 + 86 + 84 + 93 + 90}{5} \\ &= \frac{430}{5} \\ &= 86\end{aligned}$$

Complete the table.

Data Value, x	Deviation from Mean, $x - \bar{x}$	Squared Deviation, $(x - \bar{x})^2$
77	$77 - 86 = -9$	$(-9)^2 = 81$
86	$86 - 86 = 0$	$0^2 = 0$
84	$84 - 86 = -2$	$(-2)^2 = 4$
93	$93 - 86 = 7$	$7^2 = 49$
90	$90 - 86 = 4$	$4^2 = 16$

Find the mean of the squared deviations.

$$\begin{aligned}\text{mean squared deviation} &= \frac{81 + 0 + 4 + 49 + 16}{5} \\ &= \frac{150}{5} \\ &= 30\end{aligned}$$

Find the square root of the mean of the squared deviations, rounding to the nearest tenth. $\sqrt{30} = 5.5$

The standard deviation is approximately 5.5.

(B) Find the standard deviation of 3, 4, 8, 12, 7, 5, 4, 12, 3, 9, 11, 4, 14, 8, 2, 10, 3, 10, 9, 7.

Find the mean.

$$\begin{aligned}\text{mean} &= \frac{2 + 3 + 3 + 3 + 4 + 4 + 4 + 5 + 7 + 7 + 8 + 8 + 9 + 9 + 10 + 10 + 11 + 12 + 12 + 14}{20} \\ &= \frac{\boxed{}}{20} \\ &= \boxed{}\end{aligned}$$

Complete the table.

Data Value, x	Deviation from Mean, $x - \bar{x}$	Squared Deviation, $(x - \bar{x})^2$
2	$2 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
3	$3 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
3		
3		
4	$4 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
4		
4		
5	$5 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
7	$7 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
7		
8	$8 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
8		
9	$9 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
9		
10	$10 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
10		
11	$11 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
12	$12 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$
12		
14	$14 - \boxed{} = \boxed{}$	$\left(\boxed{}\right)^2 = \boxed{}$

Find the mean of the squared deviations.

$$\text{mean squared deviation} = \frac{\boxed{}}{20}$$

$$= \frac{\boxed{}}{20}$$

$$= \boxed{}$$

Find the square root of the mean of the squared deviations, rounding to the nearest tenth.

$$\sqrt{\boxed{}} = \boxed{}$$

The standard deviation is approximately _____.

Reflect

9. In terms of data values used, what makes calculating the standard deviation different from calculating the range?

_____

_____

Your Turn

10. Find the standard deviation of 21, 31, 26, 24, 28, 26.

Data Value, x	Deviation from Mean, $x - \bar{x}$	Squared Deviation, $(x - \bar{x})^2$
21		
24		
26		
26		
28		
31		

11. Find the standard deviation of 68, 71, 75, 74, 75, 71, 73, 71, 72, 74, and 79.

12. In Your Turn 11, what is the mean of the deviations before squaring? Use your answer to explain why squaring the deviations is helpful.

_____

_____

13. How can you determine the first and third quartiles of a data set?

_____

_____

_____

_____

14. How can you determine the standard deviation of a data set?

_____

_____

_____

15. **Essential Question Check-In** What does the measure of center of a data set indicate?

_____

⭐ Evaluate: Homework and Practice

- Online Homework
- Hints and Help
- Extra Practice

1. The data set $\{13, 24, 14, 15, 14\}$ gives the times of Tara's one-way ride to school (in minutes) for one week. Is the average (mean) of the times a good description of Tara's ride time? Explain.

Find the mean and median of each data set.

2. The numbers of hours Cheri works each day are 3, 7, 4, 6, and 5.

3. The weights in pounds of 6 members of a basketball team are 125, 136, 150, 119, 150, and 143.

4. 36, 18, 12, 10, 9

5. The average yearly gold price for the period from 2000–2009:

$279.11, $271.04, $309.73, $363.38, $409.72, $444.74, $603.46, $695.39, $871.96, $972.35

6. There are 28, 30, 29, 26, 31, and 30 students in a school's six Algebra 1 classes.

7. 13, 14, 18, 13, 12, 17, 15, 12

8. The numbers of members in five karate classes are 13, 12, 10, 16, and 19.

9. Find the range and interquartile range for 3, 7, 4, 6, and 5.

10. Find the range and interquartile range for 125, 136, 150, 119, 150, and 143.

11. Find the range and interquartile range for 36, 18, 12, 10, and 9.

12. Find the range and interquartile range for $279.11, $271.04, $309.73, $363.38, $409.72, $444.74, $603.46, $695.39, $871.96, and $972.35.

13. Find the range and interquartile range for 28, 30, 29, 26, 31, and 30.

14. Find the range and interquartile range for 13, 14, 18, 13, 12, 17, 15, and 12.

15. Find the range and interquartile range for 13, 12, 15, 17, and 9.

16. Find the standard deviation of 3, 7, 4, 6, and 5.

17. Find the standard deviation of 125, 136, 150, 119, 150, and 143.

18. Find the standard deviation of 36, 18, 12, 10, and 9.

19. Find the standard deviation of $279.11, $271.04, $309.73, $363.38, $409.72, $444.74, $603.46, $695.39, $871.96, and $972.35. Round the mean to the nearest $0.01 and the squared deviations to the nearest whole number.

20. Find the standard deviation of 28, 30, 29, 26, 31, and 30.

21. Find the standard deviation of 13, 14, 18, 13, 12, 17, 15, and 12.

22. Determine whether or not the third quartile has the same value as a member of the data set. Select the correct answer for each lettered part.

A. {79, 91, 90, 99, 91, 80, 80, 90}   ☐ Yes   ☐ No

B. {98, 96, 96, 95, 91, 81, 87}   ☐ Yes   ☐ No

C. {88, 95, 89, 93, 88, 93, 84, 93, 85, 92}   ☐ Yes   ☐ No

D. {97, 84, 96, 82, 93, 88, 82, 91, 94}   ☐ Yes   ☐ No

E. {94, 85, 95, 80, 97}   ☐ Yes   ☐ No

F. {85, 89, 81, 89, 85, 84}   ☐ Yes   ☐ No

Use this data for Exercises 23 and 24. The numbers of members in 6 yoga clubs are 80, 74, 77, 71, 75, and 91.

23. Find the standard deviation of the numbers of members to the nearest tenth.

H.O.T. Focus on Higher Order Thinking

24. Explain the Error Suppose a person in the club with 91 members transfers to the club with 71 members. A student claims that the measures of center and the measures of spread will all change. Correct the student's error.

25. What If? If all the values in a set are increased by 10, does the range also increase by 10? Explain.

26. Communicate Mathematical Ideas Jorge has a data set with the following values: 92, 80, 88, 95, and x. If the median value for this set is 88, what must be true about x? Explain.

27. Critical Thinking If the value for the median of a set is not found in the data set, what must be true about the data set? Explain.

Module 9

387

Lesson 1

© Houghton Mifflin Harcourt Publishing Company . Image Credits: ©Kristian Sekulic/Getty Images

Lesson Performance Task

The table lists the ages of the soprano and bass singers in a town choir. Find the mean, median, range, interquartile range, and standard deviation for each type of singer in the data set. Interpret each result. What can you conclude about the ages of the different types of singers?

Age of Soprano Singers	63	42	28	45	36	48	32	40	57	49
Age of Bass Singers	32	34	53	35	43	41	29	35	24	34

9.2 Data Distributions and Outliers

Essential Question: What statistics are most affected by outliers, and what shapes can data distributions have?

⊘ Explore   Using Dot Plots to Display Data

A **dot plot** is a data representation that uses a number line and Xs, dots, or other symbols to show frequency. Dot plots are sometimes called *line plots*.

Finance Twelve employees at a small company make the following annual salaries (in thousands of dollars): 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, and 60.

(A) Choose the number line with the most appropriate scale for this problem. Explain your reasoning.

```
←+——+——+——+——+——+——+——+——+——+——+→
  0            50           100
```

```
←+——+——+——+——+——+——+→
 20  30  40  50  60  70
```

```
←+——+——+——+——+——+——+→
 20  35  50  65  80  95
```

_____

_____

_____

_____
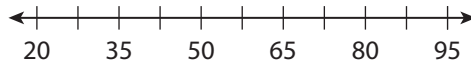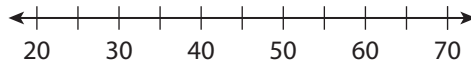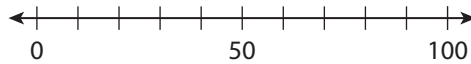
(B) Create and label a dot plot of the data. Put an X above the number line for each time that value appears in the data set.

```
←+——+——+——+——+——+——+——+——+——+——+——+→
```

Salary (thousands of dollars)

Reflect

1. **Discussion** Recall that quantitative data can be expressed as a numerical measurement. Categorical, qualitative data is expressed in categories, such as attributes or preferences. Is it appropriate to use a dot plot for displaying quantitative data, qualitative data, or both? Explain.

_____

✍ Explain 1 The Effects of an Outlier in a Data Set

An **outlier** is a value in a data set that is much greater or much less than most of the other values in the data set. Outliers are determined by using the first or third quartiles and the IQR.

How to Identify an Outlier
A data value x is an outlier if $x < Q_1 - 1.5(\text{IQR})$ or if $x > Q_3 + 1.5(\text{IQR})$.

Example 1 Create a dot plot for the data set using an appropriate scale for the number line. Determine whether the extreme value is an outlier.

(A) Suppose that the list of salaries from the Explore is expanded to include the owner's salary of $150,000. Now the list of salaries is 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, and 150.

To choose an appropriate scale, consider the minimum and maximum values, 25 and 150.

A number line from 20 to 160 will contain all the values. A scale of 5 will be convenient for the data. Label tick marks by 20s.

Plot each data value to see the distribution.

```
        X X
        X X X
    X X X X X X    X                        X
◄──┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼┼─►
    20    40    60    80   100   120   140   160
                  Salary
            (thousands of dollars)
```
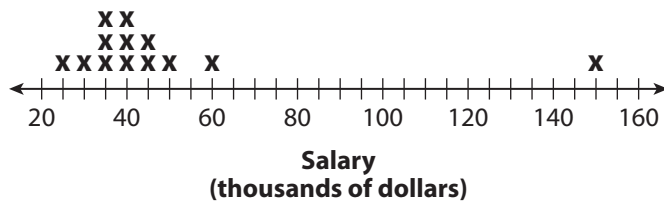
Find the quartiles and the IQR to determine whether 150 is an outlier.

```
        1-Var Stats
↑n=13
 minX=25
 Q₁=35
 Med=40
 Q₃=47.5
 maxX=150
```

$150 \overset{?}{>} Q_3 + 1.5(\text{IQR})$

$150 \overset{?}{>} 47.5 + 1.5(47.5 - 35)$

$150 > 66.25$ True

150 is an outlier.

(B) Suppose that the salaries from Part A were adjusted so that the owner's salary is $65,000.

Now the list of salaries is 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, and 65.

To choose an appropriate scale, consider the minimum and

maximum data values, _____ and _____.

A number line from _____ to _____ will

contain all the data values.

A scale of _____ will be convenient for the data.

Label tick marks by _____.

Plot each data value to see the distribution.

```
◄──┼──┼──┼──┼──┼──┼──┼──┼──┼──┼──┼──►
        Salary (thousands of dollars)
```
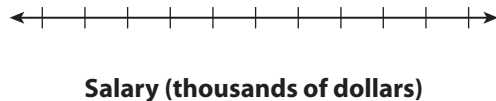
Find the quartiles and the IQR to determine whether 65 is an outlier.

```
1-Var Stats
↑n=13
 minX=25
 Q₁=35
 Med=40
 Q₃=47.5
 maxX=65
```

$65 \overset{?}{>} Q_3 + 1.5(\text{IQR})$

$65 \overset{?}{>} \boxed{} + 1.5\left(\boxed{} - \boxed{}\right)$

$65 > \boxed{}$ True / False

Therefore, 65 is / is not an outlier.

Reflect

2. Explain why the median was NOT affected by changing the max data value from 150 to 65.

_____

_____

Your Turn

3. Sports Baseball pitchers on a major league team throw at the following speeds (in miles per hour): 72, 84, 89, 81, 93, 100, 90, 88, 80, 84, and 87.

Create a dot plot using an appropriate scale for the number line. Determine whether the extreme value is an outlier.

🔑 Explain 2   Comparing Data Sets

Numbers that characterize a data set, such as measures of center and spread, are called **statistics**. They are useful when comparing large sets of data.

Example 2   **Calculate the mean, median, interquartile range (IQR), and standard deviation for each data set, and then compare the data.**

(A)   **Sports** The tables list the average ages of players on 15 teams randomly selected from the 2010 teams in the National Football League (NFL) and Major League Baseball (MLB). Describe how the average ages of NFL players compare to those of MLB players.

NFL Players' Average Ages, by Team
25.8, 26.0, 26.3, 25.7, 25.1, 25.2, 26.1, 26.4, 25.9, 26.6, 26.3, 26.2, 26.8, 25.6, 25.7

MLB Players' Average Ages, by Team
28.5, 29.0, 28.0, 27.8, 29.5, 29.1, 26.9, 28.9, 28.6, 28.7, 26.9, 30.5, 28.7, 28.9, 29.3

On a graphing calculator, enter the two sets of data into L_1 and L_2.

Use the "1-Var Stats" feature to find statistics for the data in lists L_1 and L_2. Your calculator may use the following notations: mean $\bar{x}$, standard deviation σ_x.

Scroll down to see the median (Med), Q_1, and Q_3. Complete the table.
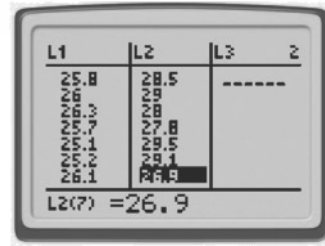
	Mean	Median	IQR ($Q_3 - Q_1$)	Standard deviation
NFL	25.98	26.00	0.60	0.46
MLB	28.62	28.70	1.10	0.91

Compare the corresponding statistics.

The mean age and median age are lower for the NFL than for the MLB, which means that NFL players tend to be younger than MLB players. In addition, the IQR and standard deviation are smaller for the NFL than for the MLB, which means that the ages of NFL players are closer together than those of MLB players.

B The tables list the ages of 10 contestants on 2 game shows.

Game Show 1
18, 20, 25, 48, 35, 39, 46, 41, 30, 27

Game Show 2
24, 29, 36, 32, 34, 41, 21, 38, 39, 26

On a graphing calculator, enter the two sets of data into L_1 and L_2.
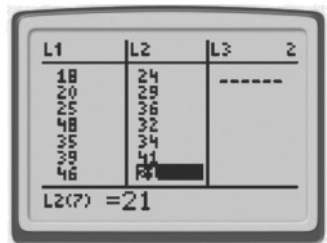
Complete the table. Then circle the correct items to compare the statistics.

	Mean	Median	IQR ($Q_3 - Q_1$)	Standard deviation
Show 1				
Show 2				

The mean is lower for the 1st / 2nd game show, which means that contestants in the 1st / 2nd game show are on average younger than contestants in the 1st / 2nd game show. However, the median is lower for the 1st / 2nd game show, which means that although contestants are on average younger on the 1st / 2nd game show, there are more young contestants on the 1st / 2nd game show. Finally, the IQR and standard deviation are higher for the 1st / 2nd game show, which means that the ages of contestants on the 1st / 2nd game show are further apart than the age of contestants on the 1st/ 2nd game show.

4. The tables list the age of each member of Congress in two randomly selected states. Complete the table and compare the data.

Illinois
26, 24, 28, 46, 39, 59, 31, 26, 64, 40, 69, 62, 31, 28, 26, 76, 57, 71, 58, 35, 32, 49, 51, 22, 33, 56

Arizona
42, 37, 58, 32, 46, 42, 26, 56, 27

	Mean	Median	IQR ($Q_3 - Q_1$)	Standard deviation
Illinois				
Arizona				

Explain 3   Comparing Data Distributions

A data distribution can be described as **symmetric**, **skewed to the left**, or **skewed to the right**, depending on the general shape of the distribution in a dot plot or other data display.
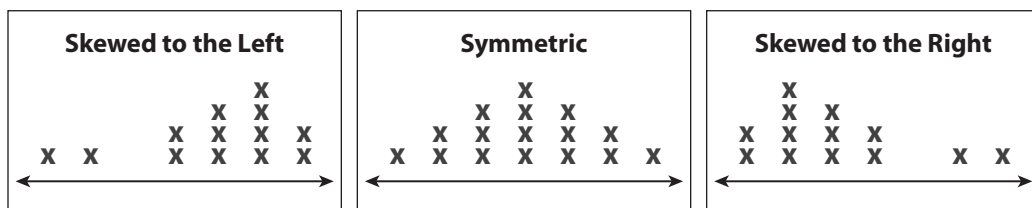
Example 3 **For each data set, make a dot plot and determine the type of distribution. Then explain what the distribution means for each data set.**

(A) **Sports** The data table shows the number of miles run by members of two track teams during one day.

Miles	3	3.5	4	4.5	5	5.5	6
Members of Team A	2	3	4	4	3	2	0
Members of Team B	1	2	2	3	3	4	3

Team A

Miles

Team B

Miles
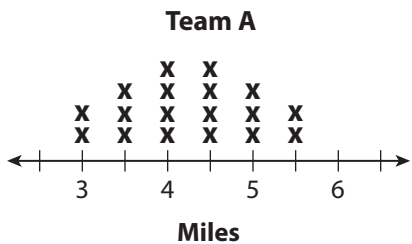
The data for team A show a symmetric distribution. This means that the distances run are evenly distributed about the mean.
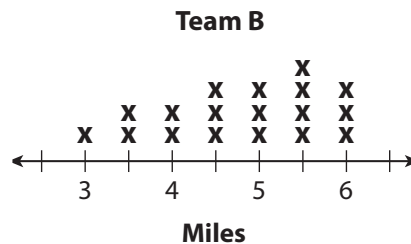
The data for team B show a distribution skewed to the left. This means that more than half the team members ran a distance greater than the mean.

B) The table shows the number of days, over the course of a month, that specific numbers of apples were sold by competing grocers.

Number of Apples Sold	0	50	100	150	200	250	300
Grocery Store A	1	4	8	8	4	1	0
Grocery Store B	3	6	8	8	2	2	1

Grocery Store A

Grocery Store B
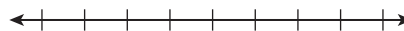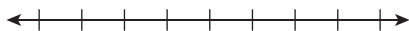
The distribution for grocery store A is: left-skewed/right-skewed / symmetric. This means that the number of apples sold each day is evenly / unevenly distributed about the mean.

The distribution for grocery store B is: left-skewed/ right-skewed /symmetric. This means that the number of apples sold each day is evenly/ unevenly distributed about the mean.

Reflect

7. Will the mean and median in a symmetric distribution always be approximately equal? Explain.

_____

_____

8. Will the mean and median in a skewed distribution always be approximately equal? Explain.

_____

_____

_____

9. **Sports** The table shows the number of free throws attempted during a basketball game. Make a dot plot and determine the type of distribution. Then explain what the distribution means for the data set.

Free Throws Shot	0	2	4	6	8
Members of Team A	2	2	4	2	2
Members of Team B	3	4	2	2	1

Team A   **Team B**

Number of Free Throws   Number of Free Throws

💬 Elaborate

10. If the mean increases after a single data point is added to a set of data, what can you tell about this data point?

11. How can you use a calculation to decide whether a data point is an outlier in a data set?

12. **Essential Question Check-In** What three shapes can data distributions have?

⭐ Evaluate: Homework and Practice

Fitness **The numbers of members in 8 workout clubs are 100, 95, 90, 85, 85, 95, 100, and 90. Use this information for Exercises 1–2.**

1. Create a dot plot for the data set using an appropriate scale for the number line.

←―┼―┼―┼―┼―┼―┼―┼―┼―┼―┼―┼―→

Number of Members

2. Suppose that a new workout club opens and immediately has 150 members. Is the number of members at this new club an outlier?

Sports **The number of feet to the left outfield wall for 10 randomly chosen baseball stadiums is 315, 325, 335, 330, 330, 330, 320, 310, 325, and 335. Use this information for Exercises 3–4.**

3. Create a dot plot for the data set using an appropriate scale for the number line.

←―┼―┼―┼―┼―┼―┼―┼―┼―┼―┼―┼―→

Number of Feet

4. The longest distance to the left outfield wall in a baseball stadium is 355 feet. Is this stadium an outlier if it is added to the data set?

Education **The numbers of students in 10 randomly chosen classes in a high school are 18, 22, 26, 31, 25, 20, 23, 26, 29, and 30. Use this information for Exercises 5–6.**

5. Create a dot plot for the data set using an appropriate scale for the number line.

←―┼―┼―┼―┼―┼―┼―┼―┼―┼―┼―┼―→

Number of Students

6. Suppose that a new class is opened for enrollment and currently has 7 students. Is this class an outlier if it is added to the data set?

Sports The average bowling scores for a group of bowlers are 200, 210, 230, 220, 230, 225, and 240. Use this information for Exercises 7–8.

7. Create a dot plot for the data set using an appropriate scale for the number line.
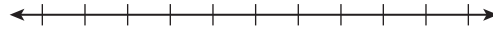
Bowling Scores

8. Suppose that a new bowler joins this group and has an average score of 275. Is this bowler an outlier in the data set?

The tables describe the average ages of employees from two randomly chosen companies. Use this information for Exercises 9–10.

Company A
23, 29, 35, 46, 51, 50, 42, 37, 30

Company B
24, 23, 45, 45, 42, 52, 55, 47, 55

9. Calculate the mean, median, interquartile range (IQR), and standard deviation for each data set.

	Mean	Median	IQR ($Q_3 - Q_1$)	Standard deviation
Company A				
Company B				

10. Compare the data sets.

The tables describe the size of microwaves, in cubic feet, chosen randomly from two competing companies. Use this information for Exercises 11–12.

Company A
1.8, 2.1, 3.1, 2.0, 3.3, 2.9, 3.3, 2.1, 3.2

Company B
1.9, 2.6, 1.8, 3.0, 2.5, 2.8, 2.0, 3.6, 3.1

11. Calculate the mean, median, interquartile range (IQR), and standard deviation for each data set.

	Mean	Median	IQR ($Q_3 - Q_1$)	Standard deviation
Company A				
Company B				

12. Compare the data sets.

For each data set, make a dot plot and determine the type of distribution. Then explain what the distribution means for each data set.

13. Sports The data table shows the number of miles run by members of two teams running a marathon.

Miles	5	10	15	20	25
Members of Team A	3	5	10	5	3
Members of Team B	6	10	4	1	5

Team A

Team B

Miles

Miles

14. Sales The data table shows the number of days that specific numbers of turkeys were sold. These days were in the two weeks before Thanksgiving.

Number of Turkeys	10	20	30	40
Grocery Store A	2	5	5	2
Grocery Store B	5	5	1	3

Grocery Store A

Grocery Store B
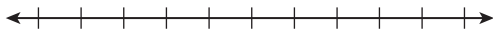
Number of Turkeys
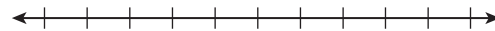
Number of Turkeys

15. State whether each set of data is left-skewed, right-skewed, or symmetrically distributed.

   A. 3, 5, 5, 3

   B. 1, 1, 3, 1

   C. 7, 9, 9, 11

   D. 5, 5, 3, 3

   E. 19, 21, 21, 19

16. What If? Given the data set 8, 15, 12, 10, and 5, what happens to the mean if you add a data value of 40? Is 40 an outlier of the new data set?

17. Critical Thinking Can an outlier be a data value between Q_1 and Q_3? Justify your answer.

18. Justify Reasoning If the distribution has outliers, why will they always have an effect on the range?

19. Education The data table describes the average testing scores in 20 randomly selected classes in two randomly selected high schools, rounded to the nearest ten. For each data set, make a dot plot, determine the type of distribution, and explain what the distribution means in context.

Average Scores	0	10	20	30	40	50	60	70	80	90	100
School A	0	1	2	2	3	4	3	2	2	1	0
School B	0	1	1	1	2	4	5	4	2	0	0

School A

School B
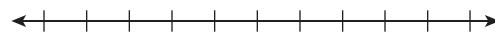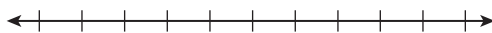
Test Scores

Test Scores

Lesson Performance Task

The tables list the daily car sales of two competing dealerships.

Dealer A			
14	13	15	12
15	16	15	17
17	12	16	14
15	16	14	16
13	14	18	15

Dealer B			
16	17	15	20
18	19	18	17
19	10	19	18
15	17	20	19
18	18	16	17

A. Calculate the mean, median, interquartile range (IQR), and standard deviation for each data set. Compare the measures of center for the two dealers.

	Mean	Median	IQR $(Q_3 - Q_1)$	Standard deviation
Dealer A				
Dealer B				

_____

_____

B. Create a dot plot for each data set. Compare the distributions of the data sets.

Dealer A

Dealer B
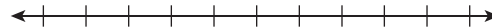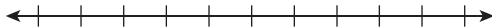
C. Determine if there are any outliers in the data sets. If there are, remove the outlier and find the statistics for that data set(s). What was affected by the outlier?

9.3 Histograms and Box Plots

Essential Question: How can you interpret and compare data sets using data displays?

⊙ Explore   Understanding Histograms

A **histogram** is a bar graph that is used to display the frequency of data divided into equal intervals. The bars must be of equal width and should touch but not overlap. The heights of the bars indicate the frequency of data values within each interval.

(A) Look at the histogram of "Scores on a Math Test." Which axis indicates the frequency?

The _____ axis shows the frequency for each interval.

(B) What does the horizontal axis indicate?

The horizontal axis shows the _____.

(C) How is the horizontal axis organized?

It is organized in groups of _____.

(D) How many had scores in the interval 60–69? ☐

(E) How many had scores in the interval 70–79? ☐

(F) How many had scores in the interval 80–89? ☐

(G) How many had scores in the interval 90–99? ☐

Score on a Math Test
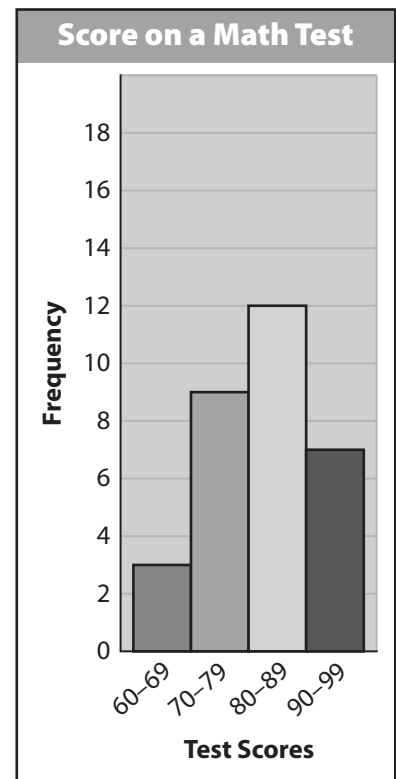
Reflect

1. What statistical information can you tell about a data set by looking at a histogram? What statistical information cannot be determined by looking at a histogram?

_____

_____

2. How many test scores were collected? How do you know?

_____

_____

🎤 Explain 1 Creating Histograms

When creating a histogram, make sure that the bars are of equal width and that they touch without overlapping. Create a frequency table to help organize the data before constructing the histogram. Consider the range of the data values when creating intervals.

Example 1 Create a frequency table from the data. Then use the frequency table to create a histogram.

(A) Listed are the ages of the 100 U.S. senators at the start of the 112th Congress on January 3, 2011.

39, 39, 42, 44, 46, 47, 47, 47, 48, 49, 49, 49, 50, 50, 51, 51,
52, 52, 53, 53, 54, 54, 55, 55, 55, 55, 55, 55, 56, 56, 57, 57,
57, 58, 58, 58, 58, 58, 59, 59, 59, 59, 60, 60, 60, 60, 60, 60,
60, 61, 61, 62, 62, 62, 63, 63, 63, 63, 64, 64, 64, 64, 66, 66,
66, 67, 67, 67, 67, 67, 67, 67, 68, 68, 68, 68, 69, 69, 69, 70,
70, 70, 71, 71, 73, 73, 74, 74, 74, 75, 76, 76, 76, 76, 77, 77,
78, 86, 86, 86

Create a frequency table. The data values range from 39 to 86, so use an interval width of 10 and start the first interval at 30.

Age Interval	Frequency
30–39	2
40–49	10
50–59	30
60–69	37
70–79	18
80–89	3

Check that the sum of the frequencies is 100.

$2 + 10 + 30 + 37 + 18 + 3 = 100$

Use the frequency table to create a histogram.
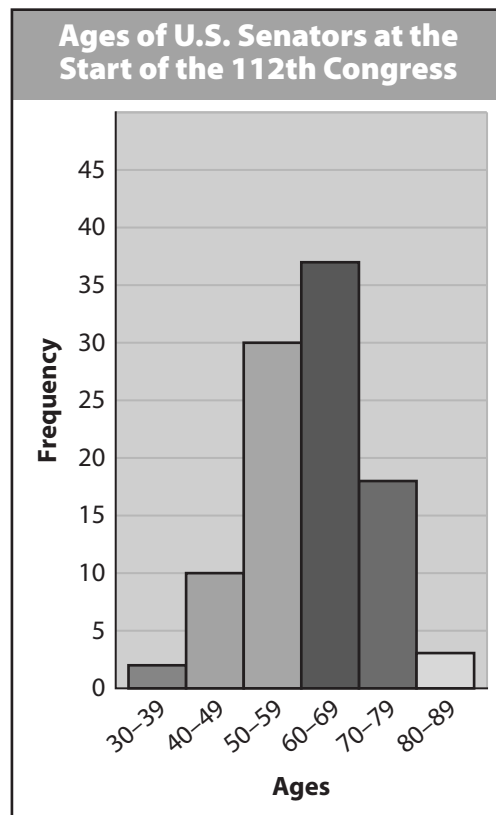
Ages of U.S. Senators at the Start of the 112th Congress

(B) Listed are the scores from a golf tournament.

68, 78, 76, 71, 69, 73, 72, 74, 76, 70, 77, 74, 75, 76, 71, 74

Create a frequency table. The data values range from [ ] to [ ],

so use an interval width of 3, and start the first interval at [ ].

Score Interval	Frequency
68–[ ]	[ ]
71–[ ]	[ ]
74–[ ]	[ ]
77–[ ]	[ ]

Check that the sum of the frequencies is [ ].

$3 + 4 + 7 + 2 =$ [ ]

Use the frequency table to create a histogram.

Golf Tournament Scores

Frequency

Scores
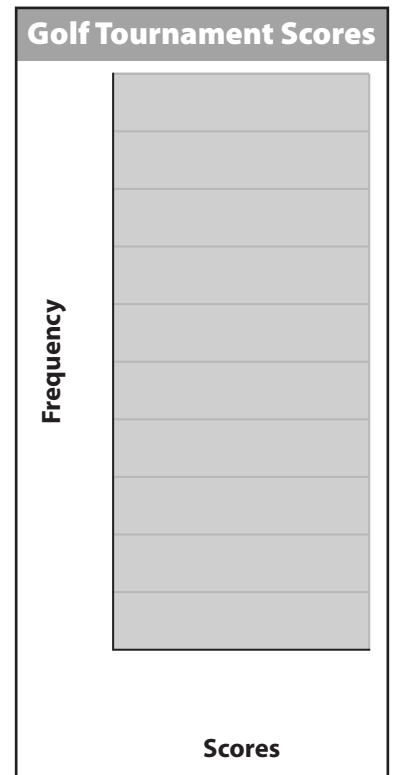
Reflect

3. Describe the shape of the distribution of senators' ages. Interpret the meaning.

_____

_____

Your Turn

4. Listed are the heights of players, in inches, on a basketball team. Create a frequency table from the data. Then use the frequency table to create a histogram.

79, 75, 74, 68, 63, 76, 74, 73, 69, 65, 71, 68, 74, 73, 70

Heights of Basketball Players

Frequency

Height
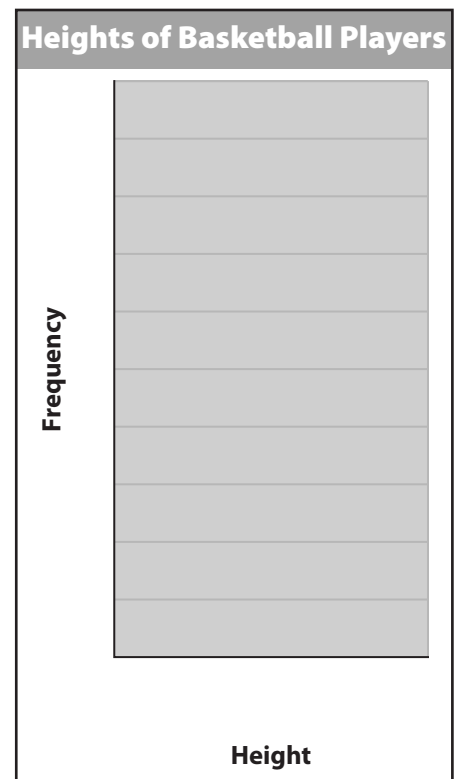
⚙ Explain 2 Estimating from Histograms

You can estimate statistics by studying a histogram.

Example 2 Estimate the mean of the data set displayed in each histogram.

(A) The histogram shows the ages of teachers in a high school.

To estimate the mean, first find the midpoint of each interval, and multiply by the frequency.

1st interval: $\left(\dfrac{20 + 29}{2}\right)(20) = (24.5)(20) = 490$

2nd interval: $\left(\dfrac{30 + 39}{2}\right)(25) = (34.5)(25) = 862.5$

3rd interval: $\left(\dfrac{40 + 49}{2}\right)(30) = (44.5)(30) = 1335$

4th interval: $\left(\dfrac{50 + 59}{2}\right)(15) = (54.5)(15) = 817.5$

5th interval: $\left(\dfrac{60 + 69}{2}\right)(10) = (64.5)(10) = 645$

Add the products and divide by the sum of the frequencies.

Mean: $\dfrac{490 + 862.5 + 1335 + 817.5 + 645}{20 + 25 + 30 + 15 + 10} = \dfrac{4150}{100} = 41.5$

A good estimate for the mean of this data set is an age of 41.5.
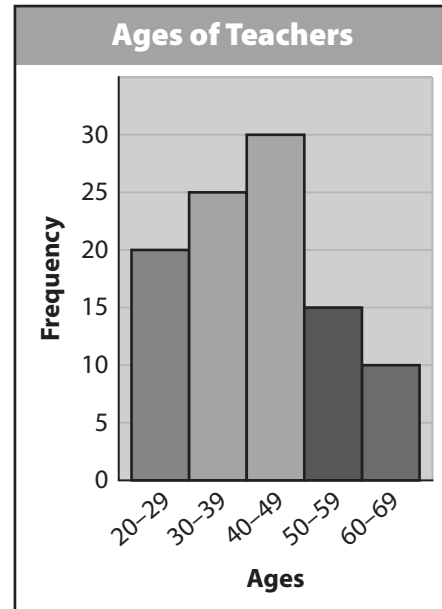
(B) The histogram shows the 2012 Olympic results for women's weightlifting.

To estimate the mean, first find the midpoint of each interval, and multiply by the frequency.

1st interval: $\left(\dfrac{160 + 179}{2}\right)\boxed{} = \boxed{}\ \boxed{} = \boxed{}$

2nd interval: $\left(\dfrac{180 + 199}{2}\right)\boxed{} = \boxed{}\ \boxed{} = \boxed{}$

3rd interval: $\left(\dfrac{200 + 219}{2}\right)\boxed{} = \boxed{}\ \boxed{} = \boxed{}$

4th interval: $\left(\dfrac{220 + 239}{2}\right)\boxed{} = \boxed{}\ \boxed{} = \boxed{}$

5th interval: $\left(\dfrac{240 + 259}{2}\right)\boxed{} = \boxed{}\ \boxed{} = \boxed{}$

Add the results and divide by the total number of values in the data set.

Mean: $\dfrac{169.5 + 379 + 838 + 2065.5 + 249.5}{1 + 2 + 4 + 9 + 1} = \dfrac{\boxed{}}{17} \approx \boxed{}$

A good estimate for the mean of this data set is $\boxed{}$ kg.
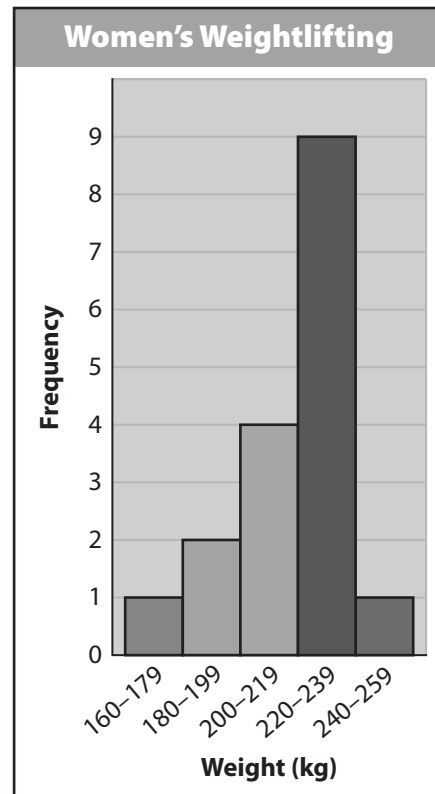
5. The histogram shows the length, in days, of Maria's last vacations. Estimate the mean of the data set displayed in the histogram.

Maria's Vacations

Length (in days)
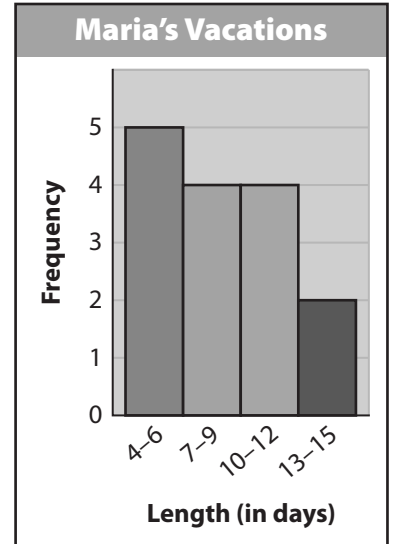
🎵 **Explain 3** **Constructing Box Plots**

A box plot can be used to show how the values in a data set are distributed. You need 5 values to make a box plot: the minimum (or least value), first quartile, median, third quartile, and maximum (or greatest value).

Example 3 **Use the data to make a box plot.**

(A) The numbers of runs scored by a softball team in 20 games are given.

3, 4, 8, 12, 7, 5, 4, 12, 3, 9, 11, 4, 14, 8, 2, 10, 3, 10, 9, 7

Order the data from least to greatest.

2, 3, 3, 3, 4, 4, 4, 5, 7, 7, 8, 8, 9, 9, 10, 10, 11, 12, 12, 14

Identify the 5 needed values. Those values are the minimum, first quartile, median, third quartile, and maximum.

$$2,\ 3,\ 3,\ 3,\ \boxed{4,\ 4,}\ 4,\ 5,\ 7,\ \boxed{7,\ 8,}\ 8,\ 9,\ 9,\ \boxed{10,\ 10,}\ 11,\ 12,\ 12,\ \boxed{14}$$

$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
Minimum	Q_1	Median	Q_3	Maximum
2	4	7.5	10	14

Draw a number line and plot a point above each of the 5 needed values. Draw a box whose ends go through the first and third quartiles, and draw a vertical line segment through the median. Draw horizontal line segments from the box to the minimum and maximum.
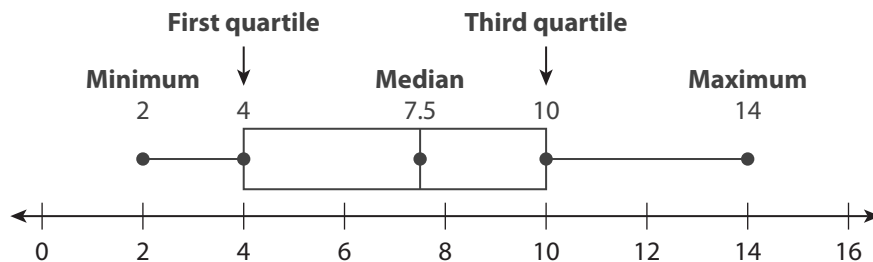
(B) 13, 14, 18, 13, 12, 17, 15, 12, 13, 19, 11, 14, 14, 18, 22, 23

Order the data from least to greatest.

☐ , 12, 12, ☐ , ☐ , 13, 14, ☐ , ☐ , 15, 17, ☐ , ☐ , 19, 22 ☐

Identify the 5 needed values. Those values are the minimum, first quartile, median, third quartile, and maximum.

☐ , 12, 12, ☐   13, 14, ☐   15, 17, ☐   19, 22, ☐

↑   ↑   ↑   ↑   ↑

Minimum Q_1 Median Q_3 Maximum

☐   ☐   ☐   ☐   ☐

Draw a number line and plot a point above each of the 5 needed values. Draw a box whose ends go through the first and third quartiles, and draw a vertical line through the median. Draw horizontal lines from the box to the minimum and maximum.

10   15   20   25
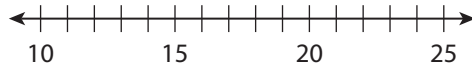
Reflect

6. The lines that extend from the box in a box plot are sometimes called "whiskers." What part (lower, middle, or upper) and about what percent of the data does the box represent? What part and about what percent does each "whisker" represent?

_____

_____

7. Which measures of spread can be determined from the box plot, and how are they found? Calculate each measure.

_____

_____

Your Turn

Use the data to make a box plot.

8. 25, 28, 26, 16, 18, 15, 25, 28, 26, 16

9. The numbers of goals scored by Lisa's soccer team in 13 games are listed below.

2, 3, 4, 1, 1, 3, 4, 2, 6, 2, 2, 3, 2

15   20   25   30

0   2   4   6   8
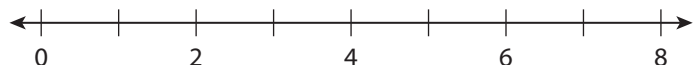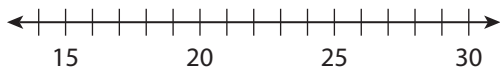
✏ Explain 4 Comparing Box Plots

You can plot two box plots above a single number line to compare two data sets.

Example 4 **Construct two box plots, one for each data set. Compare the medians and measures of variation for each distribution.**

(A) The tables show the total gross earnings, in dollars, for the top 10 movies of 2013 and 2012.

| Total Gross Earnings by the Top 25 Movies of 2013 & 2012				
Rank	Total Gross in 2013		Rank	Total Gross in 2012
1	$420,468,544		1	$623,357,910
2	$409,013,994		2	$448,139,099
3	$368,061,265		3	$408,010,692
4	$352,946,000		4	$304,360,277
5	$291,045,518		5	$303,003,568
6	$268,492,764		6	$292,324,737
7	$262,547,000		7	$262,030,663
8	$253,029,814		8	$237,283,207
9	$238,679,850		9	$218,815,487
10	$234,911,825		10	$216,391,482

🧩 Analyze Information

For each set of data, identify the five values you need to make a box plot: the

minimum, _____. In this

case, the data is from least to _____ reading from the bottom to the top.

🧩 Formulate a Plan

With the _____ needed values for each data set, construct 2 _____

on the same number line. The number line for both plots can go from $ _____

million to $ _____ million. Interpret the box plots to compare the gross
earnings for the top 10 movies in 2012 and 2013.

Solve

Using the statistics feature of a graphing calculator, find the five needed values, rounded to the nearest hundred million.

	2013 Top 10 Movies Gross Earnings ($100,000,000s)	2012 Top 10 Movies Gross Earnings ($100,000,000s)
Minimum		
Q_1		
Median		
Q_3		
Maximum		

200   250   300   350   400   450   500   550   600   650   700

The medians for both years are very close together, but gross earnings for _____ did not vary as much as in _____. The range and interquartile range were greater for the _____ data than for the _____ data. The maximum value in the 2012 data appears to be an outlier, but because _____ is not greater than _____, it is not an outlier.

Justify and Evaluate

Considering the little difference between minimum values but the great difference between maximum values the data sets, it makes sense that their measures of variation would / would not be alike.

Your Turn

Construct two box plots, one for each data set. Compare the medians and measures of variation for each distribution.

10. The net worth of the 10 richest people in the world for 2012 and 2013 (in billions) are:
2012: 69, 61, 44, 41, 37.5, 36, 30, 26, 25.5, 25.4   **2013:** 73, 67, 57, 53.5, 43, 34, 34, 31, 30, 29

20   25   30   35   40   45   50   55   60   65   70   75   80
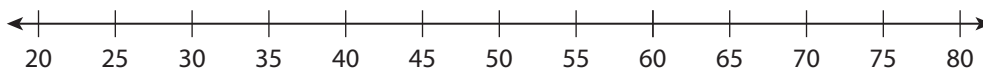
11. The ages of the 10 richest people in the world for 2012 and 2013 (in years) are:

2012: 72, 56, 81, 63, 75, 67, 55, 64, 83, 92   **2013**: 72, 57, 76, 82, 68, 77, 72, 84, 90, 63
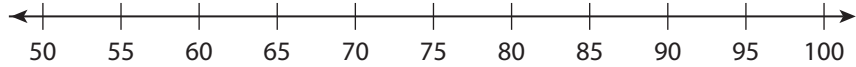
💬 **Elaborate**

12. How can you create a histogram from a data set?

_____

_____

_____

13. How can you create a box plot from a data set?

_____

_____

_____

_____

14. **Essential Question Check-In** How can you use histograms and box plots to interpret and compare data sets?

_____

_____

Use the histogram to answer the following questions.

1. What does each axis indicate?

2. How is the horizontal axis organized?

3. How many bowlers competed?

4. Describe the general shape of the distribution.

Bowlers' Scores
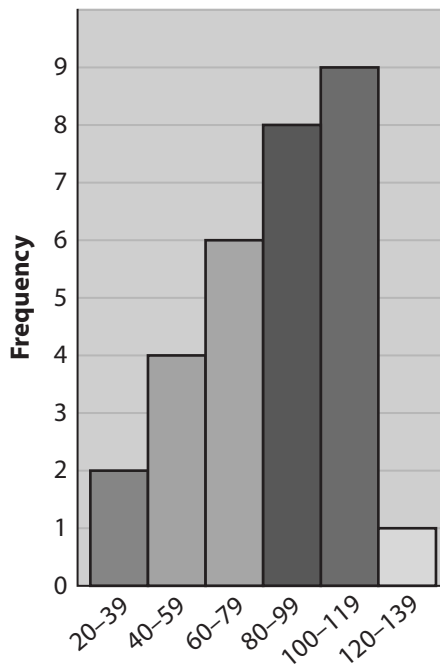
Create a histogram for the given data.

5. Listed are the ages of the first 44 U.S. presidents on the date of their first inauguration.

57, 61, 57, 57, 58, 57, 61, 54, 68, 51, 49, 64, 50, 48, 65, 52, 56, 46, 54, 49, 51, 47, 55, 55, 54, 42, 51, 56, 55, 51, 54, 51, 60, 62, 43, 55, 56, 61, 52, 69, 64, 46, 54, 47

Ages

6. Listed are the breathing intervals, in minutes, of gray whales.

8, 5, 13, 7 ,16, 9, 15, 11, 8, 6, 10, 9, 9, 11, 14, 12, 13, 15, 16, 11, 14, 9, 15, 6, 14

Time (min)

7. Listed are the heights, in inches, of the students in Marci's karate class.

42, 44, 47, 50, 51, 53, 53, 55,
56, 57, 57, 58, 59, 60, 66

Height (in.)

8. Listed are the starting salaries, in thousands of dollars, for college graduates.

34, 20, 32, 45, 32, 48, 34, 32, 20, 35, 34, 32, 40, 47, 21, 37, 21, 47, 30, 31, 40, 31, 21, 22, 30, 22, 34, 48, 35, 37, 22, 46, 38, 39, 45, 37, 52, 25, 26, 26, 27, 43, 34, 28, 55, 29, 31, 42, 24, 21, 42, 42, 31, 30, 20, 39, 23, 41, 24, 33, 49, 24, 36, 36, 23, 38, 33, 33, 54

Salary Range (thousands $)
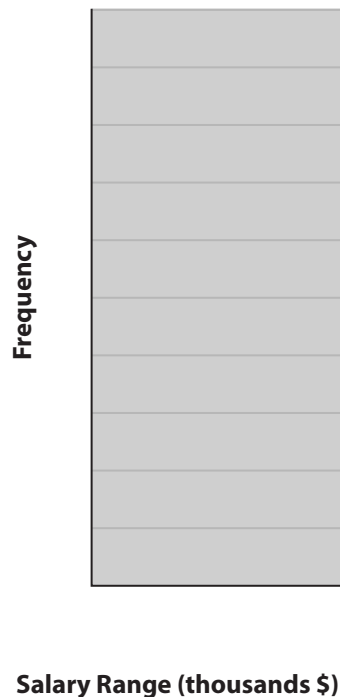
Estimate the mean of the data set displayed in each histogram.

9. The histogram shows the GPAs of the students in George's class.

GPA
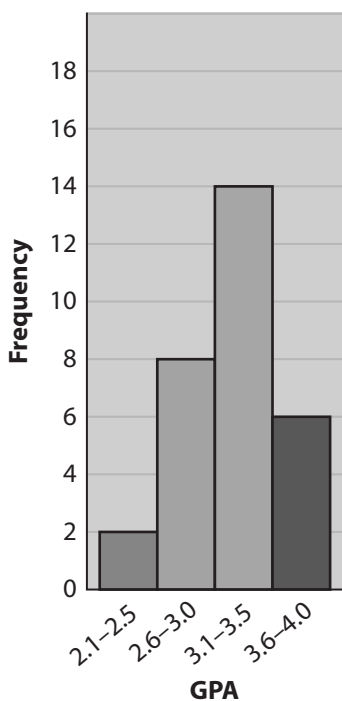
10. The histogram shows long jump distances, in feet, for a track and field team.

Frequency (y-axis: 0–9)

Distance (ft) (x-axis: 10–14, 15–19, 20–24, 25–29)
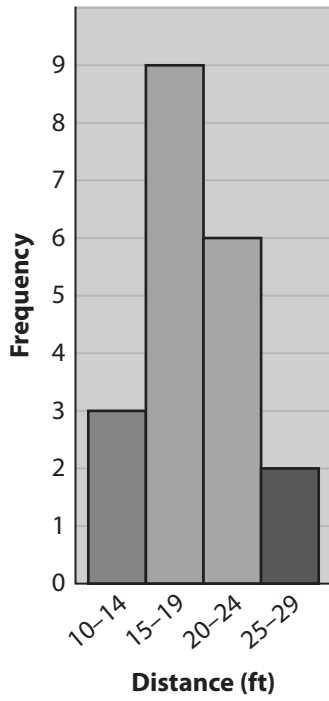
11. The histogram shows the speeds of downhill skiers, in miles per hour, during a competition.

Frequency (y-axis: 0–9)

Speed (mi/h) (x-axis: 30–39, 40–49, 50–59, 60–69, 70–79)
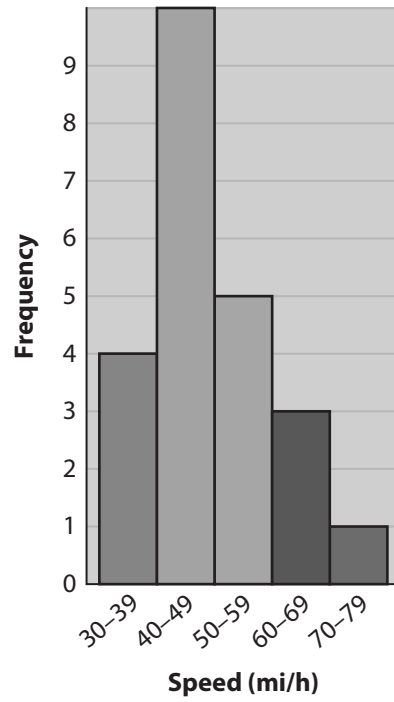
12. The histogram shows the depths a diver has been to, in meters.
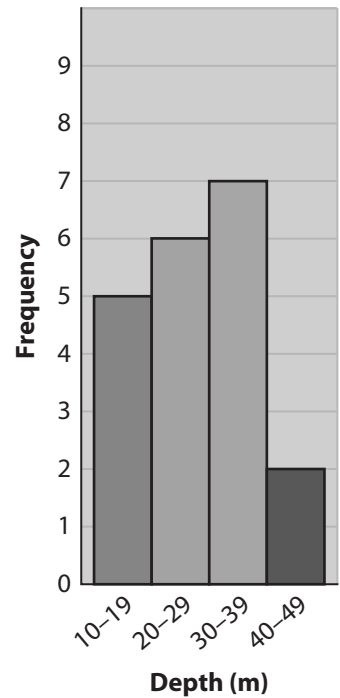
Use the data to make a box plot.

13. The numbers of points Julia's basketball team scored in 11 games are listed below.

50, 62, 37, 36, 34, 44, 44, 36, 37, 42, 36
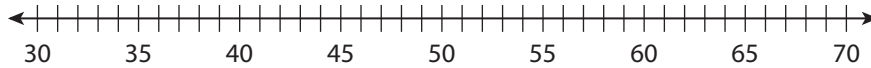
14. The numbers of baskets Kelly's team scored in 9 games are listed below.
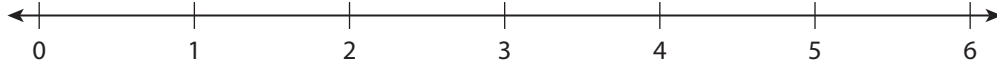
14, 24, 29, 15, 16, 20, 24, 15, 15
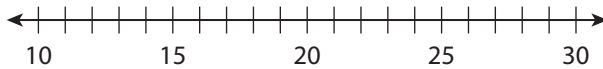
15. The numbers of runs Jane's baseball team scored in 9 games are listed below.
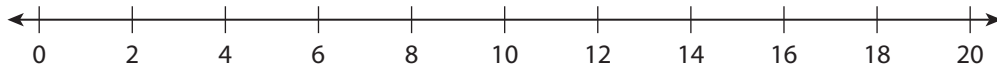
0, 2, 3, 2, 4, 11, 3, 4, 3
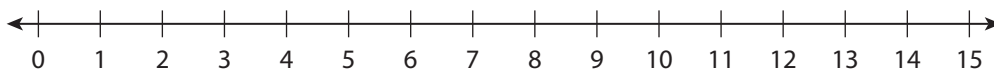
16. The numbers of field goals James' football team scored in 12 games are listed below.

3, 2, 5, 3, 5, 2, 1, 0, 5, 4, 4, 2

```
←—+———————+———————+———————+———————+———————+———————+—→
   0       1       2       3       4       5       6
```

17. The numbers of points scored by Jane in 8 basketball games are listed below.

0, 13, 12, 2, 0, 16, 3, 14

```
←—+———+———+———+———+———+———+———+———+———+———+—→
   0   2   4   6   8   10  12  14  16  18  20
```

18. The numbers of goals Claudia's soccer team scored in 21 games are shown below.

0, 5, 4, 3, 3, 2, 1, 1, 6, 2, 2, 1, 1, 1, 2, 2, 0, 1, 0, 1, 4

```
←—+———————+———————+———————+———————+———————+———————+———————+—→
   0       1       2       3       4       5       6       7
```

19. Mario and Carlos, two brothers, play for the same basketball team. Here are the points they scored in 10 games:

Game	1	2	3	4	5	6	7	8	9	10
Mario	15	X	X	12	7	11	12	11	10	11
Carlos	10	X	9	12	15	X	19	11	12	12

(Xs mark games each one missed.) Which brother had the highest-scoring game? How many more points did he score in that game than the other brother did in his highest-scoring game?

```
←—+———+———+———+———+———+———+———+———+———+———+———+———+———+———+—→
   6   7   8   9   10  11  12  13  14  15  16  17  18  19  20
```
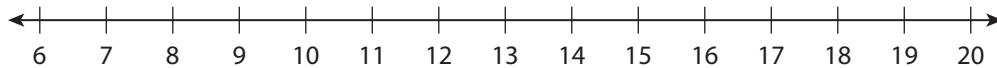
20. Communicate Mathematical Ideas Describe how you could estimate the IQR of a data set from a histogram.

21. Critical Thinking Suppose the minimum in a data set is the same as the first quartile. How would this affect a box plot of the data? Explain.

22. Draw Conclusions Dolly and Willie's scores are shown. Dolly claims that she is the better student, but Willie claims that he is the better student. What statistics make either Dolly or Willie seem like the better student? Explain.
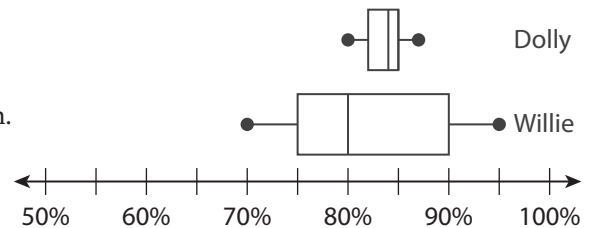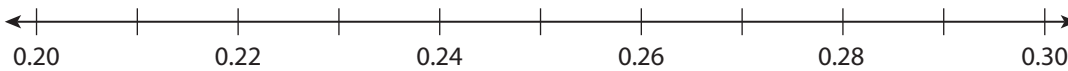
Lesson Performance Task

The batting averages for the starting lineup of two competing baseball teams are given in the table.

Batting Averages for Team A	0.270	0.260	0.300	0.290	0.260	0.280	0.280	0.240	0.270	0.280
Batting Averages for Team B	0.290	0.270	0.240	0.280	0.230	0.280	0.230	0.270	0.250	0.270

a. Create a box plot for each data set on the same number line.

Batting Averages

b. Compare the distributions of the batting averages for each team.

c. Which team has a better starting lineup? Explain.

9.4 Normal Distributions

Essential Question: How can you use characteristics of a normal distribution to make estimates and probability predictions about the population that the data represents?

⊘ Explore 1   Investigating Symmetric Distributions

A bell-shaped, symmetric distribution with a tail on each end is called a **normal distribution**.

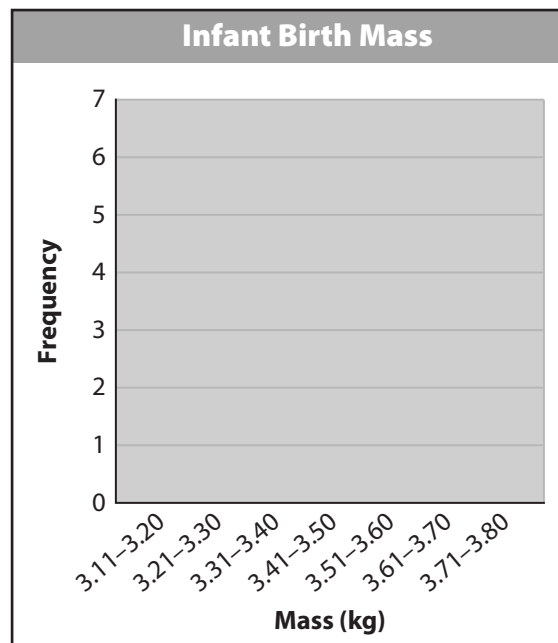Use a graphing calculator and the infant birth mass data in the table below to determine if the set represents a normal distribution.

Birth Mass (kg)				
3.3	3.6	3.5	3.4	3.7
3.6	3.5	3.4	3.7	3.5
3.4	3.5	3.2	3.6	3.4
3.8	3.5	3.6	3.3	3.5

(A) Enter the data into a graphing calculator as a list. Calculate the "1-Variable Statistics" for the distribution of data.

Mean, $\bar{x} \approx$ [ ]   Standard deviation, $\sigma_x \approx$ [ ]

Median = [ ]   IQR $= Q_3 - Q_1 =$ [ ]

(B) Sketch the histogram. Always include labels for the axes and the bar intervals.

Infant Birth Mass

(histogram with y-axis labeled Frequency from 0 to 7 and x-axis labeled Mass (kg) with intervals 3.11–3.20, 3.21–3.30, 3.31–3.40, 3.41–3.50, 3.51–3.60, 3.61–3.70, 3.71–3.80)

(C) Could this data be described by a normal distribution? Explain.

_____

_____

Reflect

1. Which intervals on the histogram had the fewest values? Which interval had the greatest number of values?

_____

_____

2. **Make a Conjecture** For the normal distribution, the mean and the median are the same. Is this true for every normal distribution? Explain.

_____

_____

_____

3. **Counterexamples** Allison thinks that every symmetric distribution must be bell-shaped. Provide a counterexample to show that she is incorrect.

_____

_____

⊘ Explore 2   **Investigating Symmetric Relative Frequency Histograms**

The table gives the frequency of each mass from the data set used in the first Explore.

Mass (kg)	3.2	3.3	3.4	3.5	3.6	3.7	3.8
Frequency	1	2	4	6	4	2	1

(A) Use the frequency table to make a relative frequency table. Notice that there are 20 data values.

Mass (kg)	3.2	3.3	3.4	3.5	3.6	3.7	3.8
Relative Frequency	$\frac{1}{20} = 0.05$						

What is the sum of the relative frequencies? _____

(B) Sketch a relative frequency histogram. The heights of the bars now indicate relative frequencies.

Infant Birth Mass

Relative frequency (y-axis): 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35

Mass (kg) (x-axis): 3.11–3.20, 3.21–3.30, 3.31–3.40, 3.41–3.50, 3.51–3.60, 3.61–3.70, 3.71–3.80
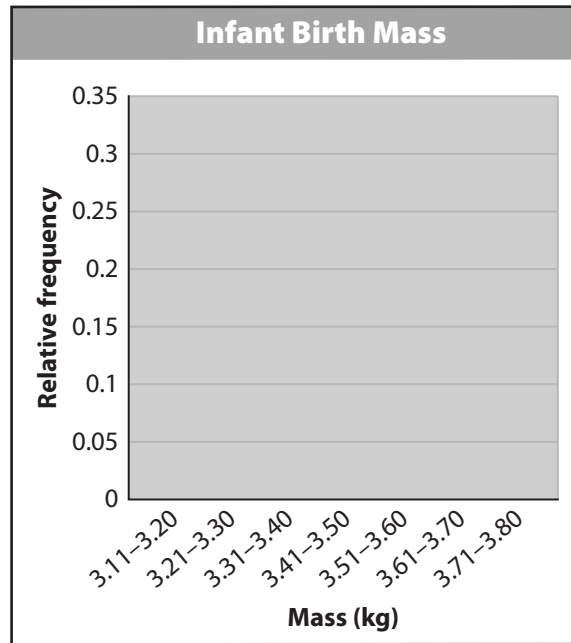
(C) Recall from the first Explore that the mean of this data set is 3.5 and the standard deviation is 0.14. By how many standard deviations does a birth mass of 3.2 kg differ from the mean? Round to one decimal place. Justify your answer.

$3.5 - 3.2 =$ ☐ and ☐ $\div\ 0.14 =$ ☐ , so a birth mass of 3.2 kg is _____

standard deviations below the mean.

Reflect

4. Identify the interval of values that are within one standard deviation of the mean. Use the frequency table to determine what percent of the values in the set are in this interval.

_____

_____

5. Identify the interval of values that are within two standard deviations of the mean. Use the frequency table to determine what percent of values in the set are in this interval.

_____

_____

© Houghton Mifflin Harcourt Publishing Company

# ⊘ Explain 1   Using Properties of Normal Distributions

The smaller the intervals are in a symmetric, bell-shaped relative frequency histogram, the closer the shape of the histogram is to a curve called a *normal curve*. Let **σ** represent the standard deviation.

A **normal curve** has the following properties:

- about 68% of the data fall within 1 standard deviation of the mean.

- about 95% of the data fall within 2 standard deviations of the mean.

- about 99.7% of the data fall within 3 standard deviations of the mean.
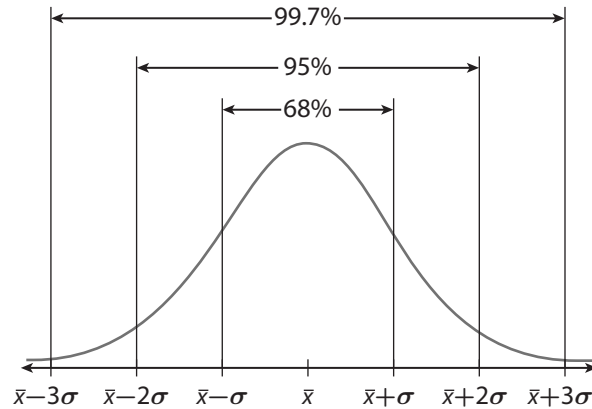
The symmetry of a normal curve allows you to separate the area under the curve into eight parts and know what percent of the data is contained in each part.
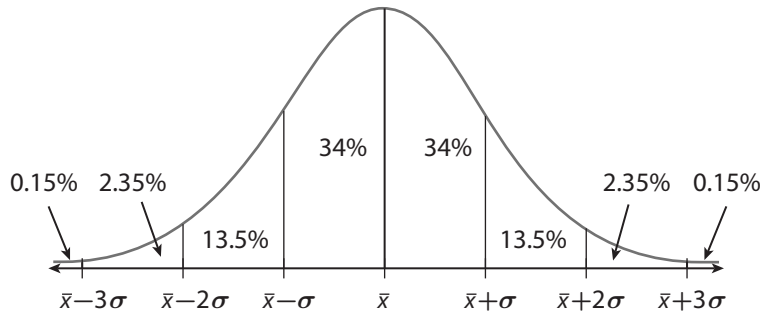
Example 1 **The masses (in grams) of pennies minted in the United States after 1982 are normally distributed with a mean of 2.50 g and a standard deviation of 0.02 g.**

(A) Find the percent of these pennies that have a mass between 2.46 g and 2.54 g.
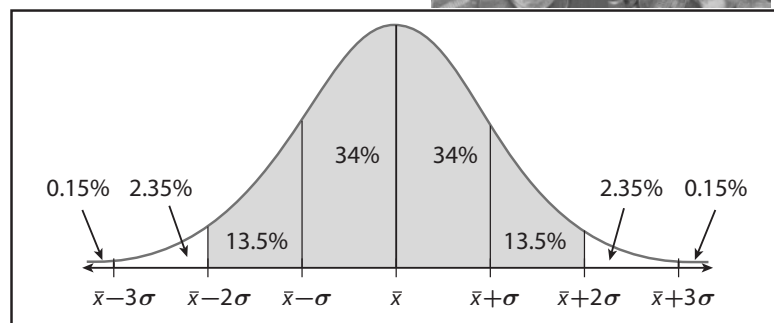
Find the percent between 2.46 g and the mean, 2.50 g.

2.50 g − 2.46 g = 0.04 g

This is 2 standard deviations below the mean.

Find the percent between 2.54 g and the mean, 2.50 g.

2.54 g − 2.50 g = 0.04 g

This is 2 standard deviations above the mean.

The shaded area represents pennies between 2.46 g and 2.54 g.

95% of the data in a normal distribution fall within 2 standard deviations of the mean.

95% of pennies minted in the United States after 1982 have a mass between 2.46 g and 2.54 g.

(B) Find the percent of pennies that have a mass between 2.48 g and 2.52 g.

Find the percent between 2.48 g and the mean, 2.50 g.

2.50 g − 2.48 g = ☐ g

This is ☐ standard deviation(s)

above/below the mean.

Shade the area for pennies between 2.48 g and 2.52 g.

Find the distance between 2.52 g and the mean, 2.50 g.

2.52 g − 2.50 g = ☐ g

This is ☐ standard deviation(s)

above/below the mean.

☐ % of the data in a normal distribution fall within ☐ standard deviation(s) of the mean.

So, ☐ % of pennies minted in the United States after 1982 have a mass between 2.48 g and 2.52 g.

Your Turn

6. Find the percent of these pennies that have a mass between 2.44 g and 2.56 g.

7. Find the percent of these pennies that have a mass between 2.46 g and 2.50 g.

🛠 Explain 2   Estimating Probabilities in Approximately Normal Distributions

You can use the properties of a normal distribution to make estimations about the larger population that the distribution represents.

Example 2   The masses (in grams) of pennies minted in the United States after 1982 are normally distributed with a mean of 2.50 g and a standard deviation of 0.02 g.

(A)   Estimate the probability that a randomly chosen penny has a mass greater than 2.52 g.
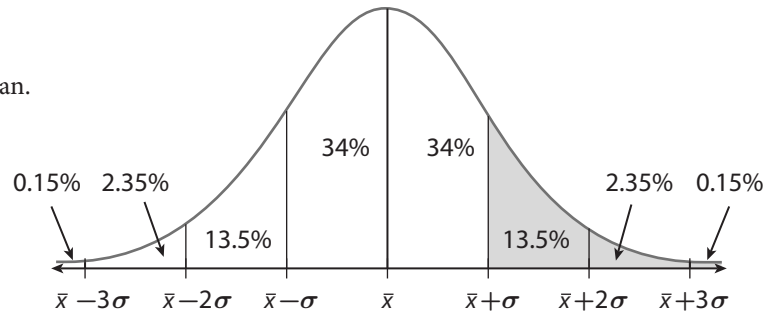
Find the percent greater than 2.52 g.
2.52 g − 2.50 g = 0.02 g
2.52 g is 1 standard deviation above the mean.

Shade in the percent of data greater than 1 standard deviation above the mean.

$13.5\% + 2.35\% + 0.15\% = 16\%$

The probability that a randomly chosen penny has a mass greater than 2.52 g is about 16%.

(B)   Estimate the probability that a randomly chosen penny has a mass greater than 2.56 g.

Find the percent greater than 2.56 g.

☐ g − ☐ g = ☐ g

☐ g is ☐ standard deviation(s) above/below the mean.
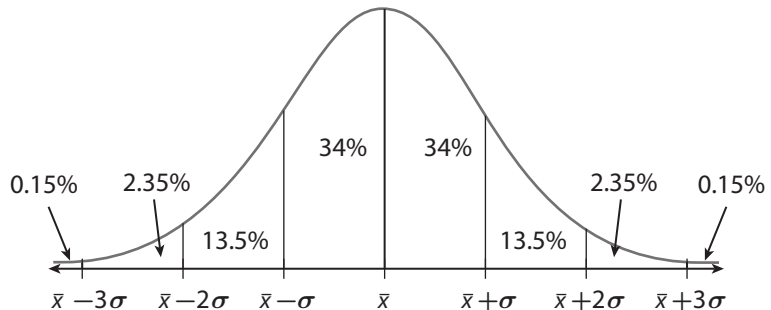
Shade in the percent of data greater/less than ☐ standard deviation(s) above/below the mean.

☐ %

The probability that a randomly chosen penny has a mass greater than 2.56 g is about ☐ %.

Your Turn

8.   Find the probability that a randomly chosen penny has a mass less than 2.54 g.

9.   Find the probability that a randomly chosen penny has a mass greater than 2.44 g.

10. For data described by a normal distribution, how do the mean and median compare?

_____

11. What does the symmetry of the normal distribution tell you about the areas above and below the mean?

_____

_____

12. How can you tell if data follow a normal distribution by looking at a histogram of the data?

_____

_____

13. **Essential Question** How do you find percents of data and probabilities of events associated with normal distributions?

_____

_____

_____

_____

⭐ Evaluate: Homework and Practice

• Online Homework
• Hints and Help
• Extra Practice

The scores on a test given to all juniors in a school district are normally distributed with a mean of 74 and a standard deviation of 8.

1. Find the percent of juniors whose score is no more than 90.

2. Find the percent of juniors whose score is between 58 and 74.

3. Find the percent of juniors whose score is at least 74.

4. Find the percent of juniors whose score is below 66.

5. Find the probability that a randomly chosen junior has a score above 82.

6. Find the probability that a randomly chosen junior has a score between 66 and 90.

7. Find the probability that a randomly chosen junior has a score below 74.

8. Find the probability that a randomly chosen junior has a score above 98.

A normal distribution has a mean of 10 and a standard deviation of 1.5.

9. Between which two values do 95% of the data fall?

10. Between which two values do 68% of the data fall?

Suppose the heights (in inches) of adult males in the United States are normally distributed with a mean of 72 inches and a standard deviation of 2 inches.

11. Find the percent of men who are no more than 68 inches tall.

12. Find the percent of men who are between 70 and 72 inches tall.

13. Find the percent of men who are at least 76 inches tall.

14. Find the probability that a randomly chosen man is more than 72 inches tall.

15. Find the probability that a randomly chosen man is between 68 and 76 inches tall.

16. Find the probability that a randomly chosen man is less than 76 inches tall.

17. Multi-Step Ten customers at Fielden Grocery were surveyed about how long they waited in line to check out. Their wait times, in minutes, are shown.

16	15	10	7	5
5	4	3	3	2

   a. What is the mean of the data set?

   b. How many data points are below the mean, and how many are above the mean?

   c. Does the data appear to be normally distributed? Explain.

18. Kori is analyzing a normal data distribution, but the data provided is incomplete. Kori knows that the mean of the data is 120 and that 84% of the data values are less than 130. Find the standard deviation for this data set.

19. Suppose compact fluorescent light bulbs last, on average, 10,000 hours. The standard deviation is 500 hours. What percent of light bulbs burn out within 11,000 hours?

20. The numbers of raisins per box in a certain brand of cereal are normally distributed with a mean of 339 raisins and a standard deviation of 9 raisins. Find the percent of boxes of this brand of cereal that have fewer than 330 raisins. Explain how you solved this problem.

Suppose the heights of professional basketball players in the United States are distributed normally, with a mean of 79 inches and a standard deviation of 4 inches.

21. How far below the mean is 71 inches and how many standard deviations is this?

22. How far below the mean is 75 inches and how many standard deviations is this?

Suppose the upper-arm length (in centimeters) of adult males in the United States is normally distributed with a mean of 39.4 cm and a standard deviation of 2.3 cm.

23. **Justify Reasoning** What percent of adult males have an upper-arm length between 34.8 cm and 41.7 cm? Explain how you got your answer.

24. **Communicate Mathematical Ideas** Explain how you can determine whether a set of data is approximately normally distributed.

25. **Critical Thinking** The distribution titled "Heads Up" shows results of many trials of tossing 6 coins and counting the number of "heads" that land facing up. The distribution titled "Number 1s Up" shows results of many trials of tossing 6 number cubes and counting the number of 1s that land facing up. For which distribution is it reasonable to use a normal distribution as an approximation? Justify your answer.

Heads Up
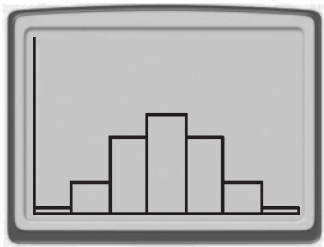
Number 1s Up
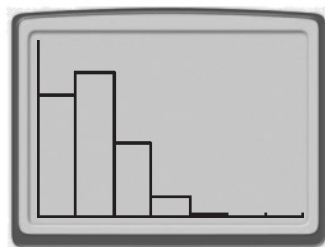
Lesson Performance Task

During a series of 100-meter races, the times of the runners are normally distributed with a mean of 12.5 seconds and a standard deviation of 0.3 seconds.

    a. Find the percent of runners that have a time between 11.6 seconds and 12.8 seconds. Explain how you got your answer.

    b. Find the probability that a randomly selected runner has a time greater than 12.8 seconds and less than 13.4 seconds.