

Chapter 10 Data Analysis

Prerequisite Skills for the Chapter “Data Analysis”

1. The *mean* of a numerical data set is the middle number when the values are written in numerical order.

2. Mean

$$\text{Sum} = 0.2 + 1.3 + 0.9 + 1.5 + 2.1 + 1.8 + 0.6 = 8.4$$

$$\text{Mean} = \frac{8.4}{7} = 1.2$$

Median

Values in order: 0.2, 0.6, 0.9, 1.3, 1.5, 1.8, 2.1

$$\text{Median} = 1.3$$

Mode

There is no mode.

3. Mean

$$\begin{aligned}\text{Sum} &= 103 + 121 + 111 + 194 + 99 + 160 + \\ &\quad 134 + 160 \\ &= 1082\end{aligned}$$

$$\text{Mean} = \frac{1082}{8} = 135.25$$

Median

Values in order: 99, 103, 111, 121, 134, 160, 160, 194

$$\text{Median} = \frac{121 + 134}{2} = \frac{255}{2} = 127.5$$

Mode

The mode is 160.

4. The distance from 0 is 2.65 units, so $|2.65| = 2.65$.
5. The distance from 0 is 1.3 units, so $|-1.3| = 1.3$.
6. The distance from 0 is 0.7 units, so $|0.7| = 0.7$.
7. The distance from 0 is 54.01 units, so $|-54.01| = 54.01$.

8. $\sqrt{289} = \sqrt{17^2} = 17$

9. $\sqrt{121} = \sqrt{11^2} = 11$

10. $\sqrt{6.25} = \sqrt{2.5^2} = 2.5$

11. $\sqrt{1.44} = \sqrt{1.2^2} = 1.2$

12. $\sqrt{80} \approx 8.94$

13. $\sqrt{13} \approx 3.61$

14. $\sqrt{25.7} \approx 5.07$

15. $\sqrt{0.89} \approx 0.94$

Lesson 10.1 Analyze Surveys and Samples

Investigating Algebra Activity for the lesson “Analyze Surveys and Samples”

1. a. Answers may vary. *Sample answer:* When the beans are stirred, the percent of beans in the handful that are red beans decreases.
b. *Sample answer:* The one handful, stirred because the beans were stirred up.
c. To get a representative mix in a sample, choose the sample randomly.

2. *Sample answer:* The two handfuls, stirred represent the beans best. Two handfuls give a larger sample, which reduces the possibility of bias from a small sample.
3. C; *Sample answer:* Mixing allows randomness and the larger reduces bias.

Guided Practice for the lesson “Analyze Surveys and Samples”

1. When all employees whose last name begins with M are surveyed, there is a rule that is followed for sampling. So, this group is a systematic sample.
2. a. Yes. People who do not attend soccer games may have significantly different opinions, so the method may result in a biased sample.
b. Yes. *Sample answer:* The question is biased because it suggests that you should have an interest in soccer because it is the world’s most popular sport. An unbiased question is “How interested are you in the sport of soccer?”

Exercises for the lesson “Analyze Surveys and Samples”

Skill Practice

1. In a *systematic* sample participants are chosen using a rule.
2. A census involves an entire population; a sample is part of a population.
3. The population is people who have dined at the restaurant. The people select themselves by mailing the comments, so this is a self-selected sampling method.
4. The population is students at your school. A rule is set for sampling, so this is a systematic sampling method.
5. The population is passengers on an airplane. The passengers are chosen at random, so this is a random sampling method.
6. A; The scientists wanted to gather information about birds.
7. The sampling is unlikely to be biased. By systematically sampling every third house on both sides of the street the family is likely to receive unbiased results.
8. The sampling method is likely to be biased. Only the nurses who access the association’s website can participate in the survey. The sample may not be representative of the nursing population.
9. The question is not potentially biased because it does not support either candidate’s tax plan.
10. The question is potentially biased because it suggests it is easy to shop online and fun to shop at the mall. An unbiased question is “Do you prefer to shop online or at the mall?”
11. The revised survey question suggests the minimum driving age is either too high or too low. An unbiased question is “Do you think the minimum driving age needs to be changed?”

12. *Sample answer:* It is possible for both manufacturers to make this claim if a biased sampling technique is used for one or both surveys.

Problem Solving

13. The question is biased because it suggests the school needs a new athletic field more than a new science lab. An unbiased question is “Do you think the school needs a new athletic field or a new science lab?”
14. The question is biased because it suggests ticket prices are already high. An unbiased question is “Would you be willing to pay more for concert tickets to finance a new arena?”
15. *Sample answer:* The sample may be biased. It does not include fans who only listen to or watch games broadcast on radio or television. The sample is also self-selected. Fans in attendance have to choose to turn in their surveys in order to be counted.
16. The samples is likely to be biased because the sampling technique for selecting the countries was not given. The 18 countries chosen might not represent the overall quality of the world’s fresh water.
17. *Sample answer:* Select every 5th student from an alphabetical list. “Do you study less than an hour each day?” The question is unbiased because it leaves two impartial answers. “Yes, I study less than an hour each day.” or “No, I study one hour or more each day.” The question is not suggesting an answer.
18. *Sample answer:* The randomized experiment; the men are similar in age and health status, and any other issues such as diet or family history should be distributed between the groups randomly. In the observational study, any occurrences of heart attack within either group might be the result of factors other than exercise, such as diet or family history.

Lesson 10.2 Use Measures of Central Tendency and Dispersion

Guided Practice for the lesson “Use Measures of Central Tendency and Dispersion”

$$1. \bar{x} = \frac{1000 + 1000 + 1181 + 1191 + 1200 + 1268 + 1328}{7} = \frac{8168}{7} = 1166.9$$

The median is the middle value, 1191.

The mode is 1000.

The median best represents the data. The mean and the mode are significantly less than most of the data.

2. a. Set A: $11.32 - 10.62 = 0.7$

400 meters: $53.68 - 46.89 = 6.79$

The range of the 400 meter set is greater than the range in set A. So, the data in the 400 meter set cover a wider interval than the data in set A.

b. The mean of set A is 11.07 and the mean absolute deviation is 0.164.

The mean of the 400 meter set is 49.82, so the mean absolute deviation is :

$$\frac{|46.89 - 49.82| + |47.65 - 49.82| + \dots + |53.68 - 49.82|}{10} = 1.725$$

The mean absolute deviation of the 400 meter set is greater, so the average variation from the mean for the data in the 400 meter set than for the data in set A.

Exercises for the lesson “Use Measures of Central Tendency and Dispersion”

Skill Practice

- The value that occurs most frequently in a data set is called the *mode* of the data.
- Measures of central tendency are used to compare typical or central values of a data set. Measures of dispersion are used to compare the spread of a data set.

3. $\bar{x} = \frac{1 + 1 + 1 + 2 + 3 + 3 + 5 + 5 + 6}{9} = \frac{27}{9} = 3$

The median is 3.

The mode is 1.

4. $\bar{x} = \frac{9 + 10 + 12 + 15 + 16}{5} = \frac{62}{5} = 12.4$

The median is 12.

There is no mode.

5. $\bar{x} = \frac{13 + 16 + 19 + 20 + 22 + 25 + 30 + 31}{8} = \frac{176}{8} = 22$

The median is $\frac{20 + 22}{2} = 21$.

There is no mode.

6. $\bar{x} = \frac{14 + 15 + 15 + 14 + 14 + 16 + 18 + 15}{8} = \frac{121}{8} = 15.125$

The median is $\frac{15 + 15}{2} = 15$.

The modes of the data are 14 and 15.

7. $\bar{x} = \frac{5.52 + 5.44 + 3.60 + 5.76 + 3.80 + 7.22}{6} = \frac{31.34}{6} \approx 5.22$

The median is $\frac{5.44 + 5.52}{2} = 5.48$.

There is no mode.

8. $\bar{x} = \frac{300 + 320 + 341 + 348 + 360 + 333}{6} = \frac{2002}{6} \approx 333.7$

The median is $\frac{333 + 341}{2} = 337$.

There is no mode.

9. B

Data in numerical order:

0.3, 0.4, 0.7, 0.7, 0.8, 0.9, 1.0, 1.2, 1.6

The median is the middle value, or 0.8.

10. The data set should be in numerical order before the median can be found.

2 3 4 4 6 6 7 8 8

The median is 6.

11. Range: $85 - 20 = 65$

The mean is 46, so the mean absolute deviation is:

$$\frac{|30 - 46| + |35 - 46| + \dots + |60 - 46|}{5} = 21.2$$

12. Range: $145 - 111 = 34$

The mean is 126, so the mean absolute deviation is:

$$\frac{|111 - 126| + |135 - 126| + \dots + |130 - 126|}{6} \approx 10.67$$

13. Range: $100 - 30 = 70$

The mean is 50.875, so the mean absolute deviation is:

$$\frac{|30 - 50.875| + |45 - 50.875| + \dots + |45 - 50.875|}{8} \approx 12.56$$

14. Range: $550 - 480 = 70$

The mean is 506.67, so the mean absolute deviation is:

$$\frac{|505 - 506.67| + |510 - 506.67| + \dots + |500 - 506.67|}{6} \approx 15.56$$

15. Range: $1.70 - 0.85 = 0.85$

The mean is about 1.26, so the mean absolute deviation is:

$$\frac{|1.25 - 1.26| + |1.50 - 1.26| + \dots + |1.25 - 1.26|}{6} \approx 0.23$$

16. Range: $84.2 - 2.5 = 81.7$

The mean is 35.52, so the mean absolute deviation is:

$$\frac{|38.2 - 35.52| + |80.1 - 35.52| + \dots + |5.5 - 35.52|}{6} \approx 31.98$$

17. The range only considers the two extreme values of a data set, while the mean absolute deviation is affected by all of the values of a data set. So, the mean absolute deviation is generally a better measure of dispersion than the range.

18. a. The mean, median, modes (if any), and range are each increased by the same constant. The mean absolute deviation is unchanged. To see why, consider adding the constant k to each value in the data set (1, 2, 3). The mean of the new data set is

$$\frac{(1+k) + (2+k) + (3+k)}{3} = \frac{1+2+3+3(k)}{3} =$$

$$\frac{1+2+3}{3} + k = \text{the original mean increased by } k.$$

The median of the new data set is $2 + k$, the original median increased by k . There is no mode, but if there were, each mode is a data value, so each would be increased by k . The greatest data value and least data value are both increased by k , so the difference of the two data values, the range, is increased by k . Each data value increases by k and the mean increases by k , so each deviation from the mean is the same as for the original data value, and the mean absolute deviation is unchanged.

- b. All five measures are multiplied by the same nonzero constant. To see why, consider multiplying each data value by the same nonzero constant k . The mean of the new data set is $\frac{k+2k+3k}{3} = \frac{6k}{3} = 2k$, which is the original mean multiplied by k . The median of the new data set is $2k$, which is the original median multiplied by k . There is no mode, but if there were, each of the modes is a data value, so each would be multiplied by k . The greatest data value and least data value are both multiplied by k , so the difference of the two data values, the range, is multiplied by k . Each data value is multiplied by k and the mean is multiplied by k , so each deviation from the mean is the absolute deviation of the original data value multiplied by k . Then the mean absolute deviation is multiplied by k .

Problem Solving

19. a. Range $24 - 5 = 19$ pounds.

b. $\bar{x} = \frac{22 + 21 + 24 + 24 + 5 + 23 + 24 + 24}{10} = \frac{196}{10} = 19.6$ pounds

The median is $\frac{23 + 24}{2} = 23.5$ pounds.

The mode is 24 lbs.

- c. The median best represents the data. The mean is lower than most of the data while the mode is higher than most of the data.

20. a. $\bar{x} = \frac{119.0 + 418.5 + \dots + 1029.9}{10} = \frac{5993.5}{10} = 599.35$ people per square mile

The median is $\frac{418.5 + 563.6}{2} = 491.05$ people per square mile.

There is no mode.

- b. The median best represents the data. There is no mode and the mean is significantly higher than most of the data.

21. a. Team 1: $202 - 150 = 52$ Team 2: $196 - 140 = 56$

The range of team 2 is greater than the range of team 1. So, the data of team 2 cover a wider interval than the data of team 1.

- b. The mean of team 1 is 171.75, so the mean absolute deviation is:

$$\frac{|162 - 171.75| + |150 - 171.75| + \dots + |202 - 171.75|}{4} = 15.75$$

The mean of team 2 is 169, so the mean absolute deviation is:

$$\frac{|140 - 169| + |153 - 169| + \dots + |196 - 169|}{4} = 22.5$$

The mean absolute deviation of team 2 is greater, so the average variation from the mean is greater for the data of team 2 than for the data of team 1.

$$22. \text{ a. } \bar{x} = \frac{230 + 80 + 65 + 40 + 25 + 15 + 11}{7} = \frac{466}{7} \approx 67$$

The median is 40.

There is no mode.

The median best represents the data. There is no mode, and the mean is greater than most of the data.

$$\text{ b. } \bar{x} = \frac{80 + 65 + 40 + 25 + 15 + 11}{6} = \frac{236}{6} \approx 39$$

The mean, in part (b) represents the data better than in part (a). In part (b) there are 3 data points greater and 3 data points less than the mean. In part (a) most of the data points are less than the mean. So, the mean in part (b) better represents the data than the mean in part (a).

$$\text{ c. Range: } 230 - 11 = 219$$

The mean is 67, so the mean absolute deviation is:

$$\frac{|230 - 67| + |80 - 67| + \dots + |11 - 67|}{7} \approx 51$$

The range is 219 and the mean absolute deviation is 51. The data lies on a large interval and the average variation from the mean is very high.

23. a. *Sample answer:* Mean: add the data values and divide by the number of data values. Median: count from left to right or from right to left to locate the middle data value or values. Mode(s): find the data value(s) with the greatest number of Xs. The mean is

$$\frac{0 + 0 + 1 + 1 + 1 + 2 + 2 + 3 + 3 + 3 + 4 + 4}{12} = 2.$$

Counting from left to right or from right to left, the two middle data values are both 2, so the median is 2. The numbers 1 and 3 have the greatest number of Xs, so the modes are 1 and 3.

- b. The weighted average of the data is

$$\frac{2 \cdot 0 + 3 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 2 \cdot 4}{12} = \frac{24}{12} = 2; \text{ whether}$$

you find the sum of the data values by simply adding them individually, or by first multiplying each distinct value by its weight and finding that sum, the result is the same. That is,

$$\frac{0 + 0 + 1 + 1 + 1 + 2 + 2 + 3 + 3 + 3 + 4 + 4}{12} =$$

$$\frac{2 \cdot 0 + 3 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 2 \cdot 4}{12} =$$

$$\frac{2}{12} \cdot 0 + \frac{3}{12} \cdot 1 + \frac{2}{12} \cdot 2 + \frac{3}{12} \cdot 3 + \frac{2}{12} \cdot 4.$$

Quiz for the lessons "Analyze Surveys and Samples" and "Use Measures of Central Tendency and Dispersion"

- The population is all hotel guests. The sampling method is self-selected sampling.
- Range: $70 - 40 = 30$

The mean is 57.5, so the mean absolute deviation is:

$$\frac{|62 - 57.5| + |63 - 57.5| + \dots + |60 - 57.5|}{6} \approx 8.33$$

The range is 30 and the mean absolute deviation is about 8.33.

- Range: $22 - 6 = 16$

The mean is 15, so the mean absolute deviation is:

$$\frac{|14 - 15| + |18 - 15| + \dots + |17 - 15|}{7} \approx 3.43$$

The range is 16 and the mean absolute deviation is about 3.43.

Extension for the lesson "Use Measures of Central Tendency and Dispersion"

$$1. \bar{x} = \frac{4 + 5 + 3 + 2 + 4 + 7 + 8 + 9 + 4 + 6 + 7 + 8 + 9 + 1}{14}$$

$$= \frac{71}{14} = 5.5$$

$$\sigma^2 = \frac{(4 - 5.5)^2 + (5 - 5.5)^2 + \dots + (1 - 5.5)^2}{14} = \frac{87.5}{14}$$

$$\approx 6.3$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{6.25} = 2.5$$

The variance is 6.3 and the standard deviation is 2.5.

$$2. \bar{x} = \frac{14 + 16 + 19 + 20 + 28 + 7 + 24 + 15 + 16 + 30 + 33 + 24}{12}$$

$$= \frac{246}{12} = 20.5$$

$$\sigma^2 = \frac{(14 - 20.5)^2 + (16 - 20.5)^2 + \dots + (24 - 20.5)^2}{12}$$

$$= \frac{625}{12} \approx 52.08$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{52.08} \approx 7.2$$

The variance is about 52.08 and the standard deviation is about 7.2.

$$3. \bar{x} = \frac{110 + 205 + 322 + 608 + 1100 + 240 + 185 + 552 + 418 + 300}{10}$$

$$= \frac{4040}{10} = 404$$

$$\sigma^2 = \frac{(110 - 404)^2 + (205 - 404)^2 + \dots + (300 - 404)^2}{10}$$

$$= \frac{766,566}{10} = 76,656.6$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{76,656.6} \approx 276.9$$

The variance is 76,656.6 and the standard deviation is about 276.9.

- The standard deviation is about 0.4.
- The standard deviation is about 29.1.
- The standard deviation is about 1192.5.
- a. Using a graphing calculator, the standard deviation is about 5.6 feet.
b. Because the height of 136 is much higher than the mean of about 80 feet, the standard deviation will increase.
c. Using a graphing calculator, the standard deviation of the new data set is about 17.7 feet. The standard deviation increased, so the prediction was correct.

Mixed Review of Problem Solving for the lessons "Analyze Surveys and Samples" and "Use Measures of Central Tendency and Dispersion"

- a. The population is patients of the doctor. Because each patient on Tuesday is asked, the sampling method is systematic sampling.
b. The survey is likely to result in a biased sample. The population sampled may not be representative

of all of the doctor's patients. The doctor should ask patients on all five days of the week which day is better to extend hours.

$$2. \bar{x} = \frac{0.25 + 0.20 + \dots + 0.10}{16} = \frac{10}{16} = 0.625 \text{ hours}$$

$$0.625 \text{ hours} \left(\frac{60 \text{ minutes}}{1 \text{ hour}} \right) = 37.5 \text{ minutes.}$$

The mean commute is 37.5 minutes.

$$3. \text{ a. Store A: } \bar{x} = \frac{280 + 200 + \dots + 190}{10} = \frac{2570}{10} = \$257$$

$$\text{The median is } \frac{230 + 260}{2} = \$245.$$

There are two modes, \$200 and \$260.

$$\text{Store B: } \bar{x} = \frac{350 + 190 + \dots + 150}{10} = \frac{2375}{10} = \$237.50$$

$$\text{The median is } \frac{200 + 220}{2} = \$210.$$

The mode is \$200.

The mean and median best represent both data sets. There is an equal number of data values on either side of the mean and median for both data sets.

b. Store A:

$$\text{Range: } 500 - 150 = \$350$$

The mean is \$257, so the mean standard deviation is :

$$\frac{|280 - 257| + |200 - 257| + \dots + |190 - 257|}{10}$$

$$= \frac{630}{10} = \$63$$

Store B:

$$\text{Range: } 400 - 150 = \$250$$

The mean is \$237.50, so the mean standard deviation is:

$$\frac{|350 - 237.5| + |190 - 237.5| + \dots + |150 - 237.5|}{10}$$

$$= \$57.50$$

The prices for Store A are more spread out than the prices for Store B. The range and mean standard deviation are both greater for Store A than Store B.

c. Store A:

Median: \$245; Lower quartile: \$200; Upper quartile: \$280; Interquartile range: $280 - 200 = \$80$. 1.5 times the interquartile range = $1.5(80) = \$120$.

An outlier for Store A is a value less than $200 - 120 = \$80$ or greater than $280 + 120 = \$400$.

Because $500 > 400$, \$500 can be considered an outlier.

Store B:

Median: \$210; Lower quartile: \$190; Upper quartile: \$250; Interquartile range: $250 - 190 = \$60$.

1.5 times the interquartile range = $1.5(60) = \$90$.

An outlier for Store B is a value less than $190 - 90 = \$100$ or greater than $250 + 90 = \$340$.

Because $350 > 340$ and $400 > 340$, \$350 and \$400 are outliers.

4. *Sample answer:* \$19, \$20, \$20, \$25, \$27, \$28, \$29, \$30, \$32, \$40

5. The students might obtain a list of registered town voters, select, say, every hundredth person on the list and contact each of them either by phone or in person.

6. Find the mean.

$$\begin{aligned} \bar{X} &= \frac{5 + 4 + 10 + 6 + 10 + 9 + 3 + 8 + 5 + 2}{10} \\ &= \frac{62}{10} = 6.2 \end{aligned}$$

Find the standard deviation.

$$\sigma = \sqrt{\frac{(5 - 6.2)^2 + (4 - 6.2)^2 + \dots + (2 - 6.2)^2}{10}} \approx 3.1$$

The standard deviation of the times is 3.1 minutes.

Lesson 10.3 Analyzing Data

Guided Practice for the lesson "Analyzing Data"

- In Mr. Cole's homeroom, 9 students prefer English while 7 students prefer science. More students prefer English.
- The total number of students playing soccer is $152 - (22 + 50) = 80$. The number of girls playing golf is $22 - 10 = 12$. The number of boys playing basketball is $50 - 26 = 24$. The number of boys playing soccer is $80 - (10 + 24) = 46$. The number of girls playing soccer is $80 - 46$.

	Golf	Basketball	Soccer	Total
Boys	10	24	46	80
Girls	12	26	34	72
Total	22	50	80	152

Exercises for the lesson "Analyzing Data"

Skill Practice

- The body of a two-way frequency table gives the *joint frequency* of the categories involved.
- Find the total of a row or column. *Sample answer:* In the table for Exercises 3–5, the marginal frequency of labradors is 11.
- The cell where the Males row and the Poodles column cross contains the number 5, so the pet store has 5 male poodles.
- The cell where the Females row and the Total column cross contains the number 18, so the pet store has 18 female puppies.
- There are 11 labrador puppies and 9 yorkie puppies. So the store has $11 - 9$, or 2 more labrador puppies than yorkie puppies.
- The total number of 2-door cars is $2465 - 1564 = 901$. The number of 8-cylinder 2-door cars is then $901 - 586 = 315$. The number of 6-cylinder 4-door cars is $1564 - 840 = 724$. So the total number of 6-cylinder cars is $586 + 724 = 1310$, and the total number of 8-cylinder cars is $315 + 840 = 1155$.

	2 door	4 door	Total
6 cylinder	586	724	1310
8 cylinder	315	840	1155
Total	901	1564	2465

7. Begin by creating a two-way table with 0 Sisters, 1 Sister, 2 Sisters, and Total along the top and 0 Brothers, 1 Brother, 2 Brothers, and Total along the side. Fill in the portions of the table using the given information.

- There are 82 total students.
- 23 students have 2 brothers.
- 28 students have 1 brother.
- 9 students have no brothers or sisters.
- 10 students have only 1 sister (no brothers).
- 7 students have only 1 brother (no sisters).
- 6 students have 2 sisters and 1 brother.
- 22 students have 2 sisters.
- 27 students have 0 sisters.
- 8 students have 1 sister and 2 brothers.

	0 Sisters	1 Sister	2 Sisters	Total
0 Brothers	9	10		
1 Brother	7		6	28
2 Brothers		8		23
Total	27		22	82

Now complete the table. The total number of students with 0 brothers is $82 - (28 + 23)$, or 31 students. The total number of students with 1 sister is $82 - (27 + 22)$, or 33 students. The number of students with 2 brothers and 0 sisters is $27 - (9 + 7)$, or 11 students. The number of students with 1 brother and 1 sister is $28 - (7 + 6)$, or 15 students. Then the number of students with 2 brothers and 2 sisters is $23 - (11 + 8)$, or 4 students. And the number of students with 0 brothers and 2 sisters is $31 - (9 + 10)$, or 12 students.

	0 Sisters	1 Sister	2 Sisters	Total
0 Brothers	9	10	12	31
1 Brother	7	15	6	28
2 Brothers	11	8	4	23
Total	27	33	22	82

8. The value where the 4 Bedroom column and the 3 Bath row cross is 75. The correct answer choice is D.
9. Matt received more votes from the juniors than either of the other two candidates, so Matt must be a junior.
10. Yes; Katy had the most votes from both the freshmen class and the senior class.
11. Katy won the election; she received a total of 448 votes, more than both Matt's 359 votes and Olivia's 326 votes.

12. Begin by creating a two-way table with 12 oz, 16 oz, and Total along the top and Water, Iced Tea, and Total along the side. Fill in the portions of the table using the given information.

- The number of 16-ounce bottles is one less than the number of 12-ounce bottles.
- There are 11 more bottles of iced tea than water.
- There are 16 bottles of water.
- The number of 12-ounce bottles of water is two less than twice the number of 16-ounce bottles of water.

Using the first piece of information, let x be the total number of 12-ounce bottles and then $x - 1$ will represent total number of 16-ounce bottles. Using the last piece of information, let y be the number of 16-ounce bottles of water and then $2y - 2$ will represent total number of 12-ounce bottles of water.

	12 ounce	16 ounce	Total
Water	$2y - 2$	y	16
Iced Tea			
Total	x	$x - 1$	

Because there are 16 bottles of water and the second piece of information states that there are 11 more bottles of iced tea than water, there are $11 + 16$, or 27 bottles of iced tea. Therefore, there is a total of $16 + 27$, or 43 bottles in all.

	12 oz	16 oz	Total
Water	$2y - 2$	y	16
Iced Tea			27
Total	x	$x - 1$	43

Now find the values of x and y .

$$\begin{aligned} x + (x - 1) &= 43 & (2y - 2) + y &= 16 \\ 2x - 1 &= 43 & 3y - 2 &= 16 \\ 2x &= 44 & 3y &= 18 \\ x &= 22 & y &= 6 \end{aligned}$$

Because $x = 22$, the total number of 12-ounce bottles is 22 and the number of 16-ounce bottles is $22 - 1$, or 21. Because $y = 6$, the number of 16-ounce bottles of water is 6 and the number of 12-ounce bottles of water is $2(6) - 2$, or 10. Enter the values in the table.

	12 ounce	16 ounce	Total
Water	10	6	16
Iced Tea			27
Total	22	21	43

Because the total number of 12-ounce bottles is 22 and 10 of them contain water, the number of 12-ounce bottles of iced tea is $22 - 10$, or 12. Because the total number of 16-ounce bottles is 21 and 6 of them contain water, the number of 16-ounce bottles of iced tea is $21 - 6$, or 15. Complete the table.

	12 ounce	16 ounce	Total
Water	10	6	16
Iced Tea	12	15	27
Total	22	21	43

Problem Solving

13. There were 97 ham sandwiches on wheat bread sold at the deli and 41 chicken sandwiches on white bread. There were $97 - 41$, or 56 more ham sandwiches on wheat bread sold than chicken sandwiches on white bread.
14. Chicken on wheat bread. *Sample answer:* There were 75 chicken sandwiches on wheat bread sold and only 65 ham sandwiches on white bread sold. So if I chose a sandwich at random, it would more likely be a chicken sandwich on wheat bread because more of this kind was sold.
15. Ham on wheat bread. *Sample answer:* Looking at the Wheat bread row in the table, the greatest value in that row is 97, in the Ham column. So if I knew that a customer was going to order a sandwich on wheat bread, the customer's most likely choice is a ham sandwich.
16. a. Because there are 33 students in choir, of which 6 are less than 5 feet tall and 22 are between 5 and 6 feet tall, there are $33 - (6 + 22)$, or 5 choir members who are more than 6 feet tall.
- b. The total number of choir and band members is $33 + 74$, or 107 students. Of these 107 students, 23 are less than 5 feet tall and 24 are more than 6 feet tall, so there are $107 - (23 + 24)$, or 60 students who are between 5 and 6 feet tall. Of these 60 students, 22 are choir members, so there are $60 - 22$, or 38 band members who are between 5 and 6 feet tall.
- c. Choir member. *Sample answer:* Of the 33 students in choir, 22 of them are between 5 and 6 feet tall or about 67%. Of the 74 students in band, 38 of them are between 5 and 6 feet tall or about 51%. So a student selected at random from the choir is more likely to be between 5 and 6 feet tall than a student selected at random from the band.
17. a. Yes, fertilizer-fortified soil produced 56 tomatoes while soil fertilized every 2 weeks produced 65 tomatoes. So, soil fertilized every 2 weeks is the better treatment for tomatoes.
- b. Yes, fertilizer-fortified soil produced 37 green peppers while soil fertilized every 2 weeks produced 19 green peppers. So, fertilizer-fortified soil is the better treatment for green peppers.
- c. Fertilizer-fortified soil appears to be the better choice since it produced 93 total vegetables while soil fertilized every 2 weeks only produced 84 vegetables; no, fertilizer-fortified soil is the better choice for green peppers, but it is not the better choice for tomatoes.
18. a. The total number of rock CDs is $141 - 47$, or 94. So the number of rock CDs that Maleho owns is $94 - (19 + 38)$, or 37 and the number of classical CDs Maleho owns is $49 - 37$, or 12. The total number of

CDs owned by Sangee is $26 + 19$, or 45. So the total number of CDs owned by Tom is $141 - (45 + 49)$, or 47. Finally, the number of classical CDs that Tom owns is $47 - (26 + 12)$, or 9.

	Classical	Rock	Total
Sangee	26	19	45
Tom	9	38	47
Maleho	12	37	49
Total	47	94	141

- b. After the purchase, Sangee would have 27 classical CDs. Because Maleho has 37 rock CDs, Sangee's new classical CD total would be $37 - 27$, or 10 less than Maleho's rock CD total.
- c. Since the boys have a total of 94 rock CDs but only 47 classical CDs, if a CD is chosen at random it would be more likely for it to be a rock CD.
19. a. Since 1809 tickets were sold, of which 800 tickets were for the main floor, there were $1809 - 800 = 1009$ tickets sold for the balcony.
- Write an equation for the balcony tickets:
- $$(2x + y) + (x - 40) = 800$$
- $$3x + y - 40 = 800$$
- $$3x + y = 840$$
- Write an equation for the main floor tickets:
- $$(x + 2y) + (3x - y - 80) = 1009$$
- $$4x + y - 80 = 1009$$
- $$4x + y = 1089$$
- Now solve the system for x and y . Start by subtracting the first equation from the second equation.
- $$4x + y = 1089$$
- $$\underline{3x + y = 840}$$
- $$x = 249$$
- Replace x with 249 in either equation and solve for y .
- $$3x + y = 840$$
- $$3(249) + y = 840$$
- $$747 + y = 840$$
- $$y = 93$$
- So, $x = 249$ and $y = 93$.
- b. $x + 2y = 249 + 2(93) = 435$ tickets
- c. $x - 40 = 249 - 40 = 209$ tickets

Lesson 10.4 Interpret Stem-and-Leaf Plots and Histograms

Investigating Algebra Activity for the lesson "Interpret Stem-and-Leaf Plots and Histograms"

- 1–4. Answers will vary.
5. *Sample answer:* The temperatures would be higher and would not include negative values.
6. *Sample answer:* The data in Plot A are closer together than the data in Plot B. The values in Plot A are more predictable.

Guided Practice for the lesson “Interpret Stem-and-Leaf Plots and Histograms”

1. Years

Stem	Leaves
178	8 7 8 8 8 8 7 8 9 7 8 8
179	2 0 6 1
180	3
181	6 2 7

Key: 178 | 8 = Year 1788

Years

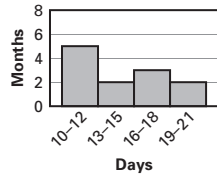
Stem	Leaves
178	7 7 7 8 8 8 8 8 8 8 8 9
179	0 1 2 6
180	3
181	2 6 7

Key: 178 | 8 = Year 1788

- The intervals for 0–9 home runs and 10–19 home runs contain the majority of the data. So, most of the baseball players hit between 3 and 19 home runs.
- Consider the distribution of the data. The intervals from 10–29% contain most of the data for the female students. The intervals from 20–39% contain most of the data for the males. The clustering of data shows that the male students generally watch more television each day than female students.

Days monthly	Months
10–12	
13–15	
16–18	
19–21	

Rainy Days in Buffalo, NY



Exercises for the lesson “Interpret Stem-and-Leaf Plots and Histograms”

Skill Practice

- The number of data values in an interval is the *frequency* of that interval.
- A histogram displays numerical data from a frequency table. A bar graph displays categorical data.

Stem	Leaves
1	7
2	0 0 4
3	0 1 3 5 8 8 9
4	2
5	5

Key: 4 | 2 = 42

Stem	Leaves
0	2 5 6 6 7 8 8 9
1	1 4 7
2	0
3	2 3

Key: 1 | 7 = 17

Stem	Leaves
10	5 9
11	1
12	1 4 7
13	3
14	2
15	6
16	
17	9
18	2

Key: 13 | 3 = 133

Stem	Leaves
1.0	5
1.1	1 9 9
1.2	2 3 9
1.3	1 5
1.4	5

Key: 1.3 | 1 = 1.31

- The stem-and-leaf plot should have a key that tells how to read the data.

1	8 8 8 9 9 9
2	0 0 0 1 1 1 2 3

Key: 1 | 8 = 18

- C;

The median is the middle value of the data set:

21, 21, 21, 32, 33, 33, 34, 40, 42.

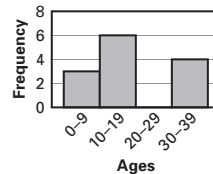
The median is 33.

- C;

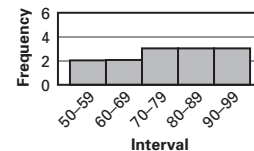
Range = 51 – 31 = 20

The range is 20.

- The interval 20–29 has a frequency of zero but should be displayed on the histogram. There needs to be a space the same width as the other bars that indicates the frequency of the interval 20–29 is zero.

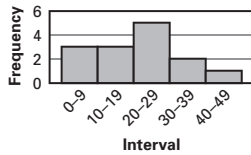


Interval	Frequency
50–59	
60–69	
70–79	
80–89	
90–99	



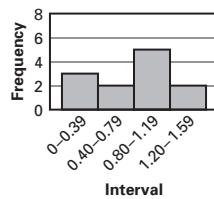
12.

Interval	Frequency
0–9	III
10–19	III
20–29	IIII
30–39	II
40–49	I



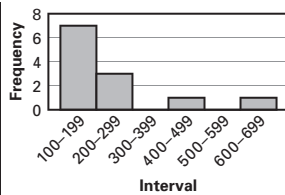
13.

Interval	Frequency
0–0.39	III
0.40–0.79	II
0.80–1.19	IIII
1.20–1.59	II



14.

Interval	Frequency
100–199	IIII II
200–299	III
300–399	
400–499	I
500–599	
600–699	I



15. A stem-and-leaf plot of the data contains only 2 intervals; 10–19 and 20–29. The plot would indicate that all but two data values are in the 10–19 interval. Unlike a stem-and-leaf plot, a histogram can be broken down into smaller intervals. This allows the distribution of the data to be examined closer.

16.

Interval	0–8	9–17	18–26	27–35
Frequency	III	IIII	IIII	II

Sample answer: The plot needs to contain 3 values in the range 0–8, 5 in the range 9–17, 4 in the range 18–26, and 2 in the range 27–35. A correct solution will contain the proper number of values in each interval.

Stem		Leaves
0		1 3 5 9
1		2 2 4 4 8 9
2		0 2 7 8

Key: 1 | 2 = 12

Problem Solving

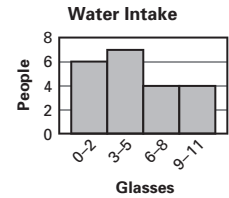
17. *Heights*

Stem		Leaves
6		8 9 9
7		0 2 4 4 6 8 8
8		0 1 1 1 1

Key: 7 | 2 = 72 inches

18. *Glasses* | People

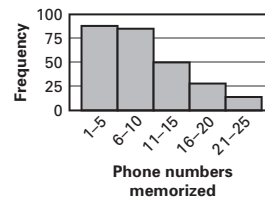
0–2		IIII I
3–5		IIII II
6–8		IIII
9–11		IIII



19. a.

Phone numbers	1–5	6–10	11–15	16–20	21–25
Frequency	88	85	50	28	14

Phone Number Memorization



b. $P(11–25 \text{ numbers})$

$$= \frac{\text{Number of people with 11–25 phone numbers}}{\text{Total number of people surveyed}}$$

$$= \frac{50 + 28 + 14}{88 + 85 + 50 + 28 + 14} = \frac{92}{265} \approx 0.35$$

The probability that a person surveyed has 11–25 phone numbers memorized is $\frac{92}{265}$ or about 0.35.

20. a. *House*: median = 132 days

Senate: median = 149 days

b. *House*: range = 142 – 110 = 32 days

Senate: range = 173 – 132 = 41 days

c. The majority of the data for the House is below 140 days while the majority of the data of the Senate is above 140 days. This distribution shows the Senate is generally in session for more days a year than the House of Representatives.

21. a. *Ages*

Stem		Leaves
1		8
2		0 0 0 1 1 1 1 1 5 5 5 9
3		0 0 0 2 2 4 4 5 7 8 8 8 8 8 9
4		0 1 2 5 5 8 9
5		0 4 5
6		4

Key: 1 | 8 = 18 years

b. Median age = 34 years

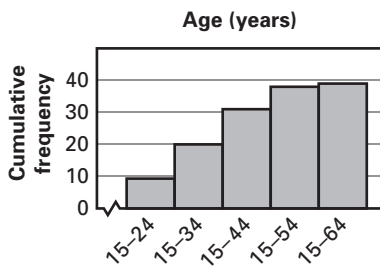
$$\text{Range } 64 - 18 = 46 \text{ years}$$

$$\begin{aligned} \text{c. } P(18-29 \text{ years}) &= \frac{\text{Number of males } 18-29 \text{ years}}{\text{Total number of males}} \\ &= \frac{1 + 12}{39} = \frac{13}{39} = \frac{1}{3} \end{aligned}$$

The probability Thomas English was 18–29 years old is $\frac{1}{3}$.

22.

Ages	Frequency	Cumulative Frequency
15-24	9	9
25-34	11	9 + 11 = 20
35-44	11	20 + 11 = 31
45-54	6	31 + 6 = 37
55-64	2	37 + 2 = 39



Lesson 10.5 Interpret Box-and-Whisker Plots

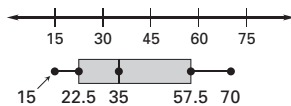
Guided Practice for the lesson "Interpret Box-and-Whisker Plots"

1. Ordered data: 15, 20, 25, 30, 40, 55, 60, 70

$$\text{The median is } \frac{30 + 40}{2} = 35.$$

$$\text{The lower quartile is } \frac{20 + 25}{2} = 22.5.$$

$$\text{The upper quartile is } \frac{55 + 60}{2} = 57.5.$$



2. For Dallas, the median precipitation is 2.6 inches. Above the median is 50% of the data, so for 50% of 12 months, or 6 months, Dallas has more than 2.6 inches of precipitation.

3. D

Ordered data: 3.0, 3.2, 3.4, 3.6, 3.7, 3.7, 3.8, 4.2, 4.3, 4.5, 5.2, 5.4

$$\text{The median is } \frac{3.7 + 3.8}{2} = 3.75.$$

$$\text{The lower quartile is } \frac{3.4 + 3.6}{2} = 3.5.$$

$$\text{The upper quartile is } \frac{4.3 + 4.5}{2} = 4.4.$$

The interquartile range is $4.4 - 3.5 = 1.1$.

Find 1.5 times the interquartile range: $1.5(1.1) = 1.65$.

A value that is less than $3.5 - 1.65 = 1.85$ is an outlier.

A value that is greater than $4.4 + 1.65 = 6.05$ is an outlier.

So, there are no outliers in this data set.

Exercises for the lesson "Interpret Box-and-Whisker Plots"

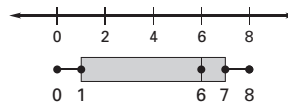
Skill Practice

- The interquartile range of a data set is the difference of the upper quartile and the lower quartile.
- If a value in a data set is greater than the upper quartile by more than 1.5 times the interquartile range or less than the lower quartile by more than 1.5 times the interquartile range, the data value is an outlier.
- Ordered data: 0, 0, 1, 2, 3, 6, 6, 7, 7, 7, 8

The median is 6.

The lower quartile is 1.

The upper quartile is 7.

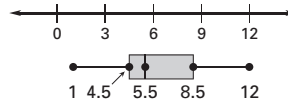


4. Ordered data: 1, 1, 4, 5, 5, 5, 6, 7, 8, 9, 10, 12

$$\text{The median is } \frac{5 + 6}{2} = 5.5.$$

$$\text{The lower quartile is } \frac{4 + 5}{2} = 4.5.$$

$$\text{The upper quartile is } \frac{8 + 9}{2} = 8.5.$$

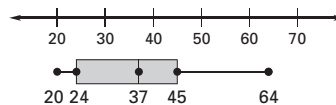


5. Ordered data: 20, 23, 24, 32, 35, 39, 42, 45, 52, 64

$$\text{The median is } \frac{35 + 39}{2} = 37.$$

The lower quartile is 24.

The upper quartile is 45.

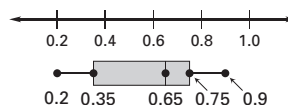


6. Ordered data: 0.2, 0.3, 0.4, 0.6, 0.7, 0.7, 0.8, 0.9

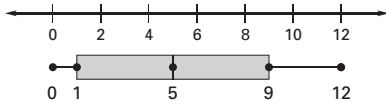
$$\text{The median is } \frac{0.6 + 0.7}{2} = 0.65.$$

$$\text{The lower quartile is } \frac{0.3 + 0.4}{2} = 0.35.$$

$$\text{The upper quartile is } \frac{0.7 + 0.8}{2} = 0.75.$$

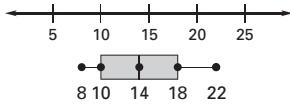


7. The upper quartile is incorrect. The upper quartile of the data should be $\frac{8 + 10}{2} = 9$.



8. A; A whisker represents 25% of the data. Since 20 is the upper quartile of the data, 25% of the data is greater than 20.
9. B; The median of the data is 15. Below the median is 50% of the data.
10. The percentage is incorrect. The box on each side of the median represents 25% of the data, so about 50% of the data values lie between 11 and 20.
11. Ordered values: 8, 10, 10, 10, 11, 12, 12, 14, 15, 15, 15, 18, 18, 22, 22

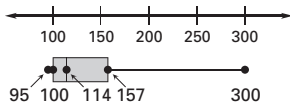
The median is 14 hours.
 The lower quartile is 10 hours.
 The upper quartile is 18 hours.
 The interquartile range is $18 - 10 = 8$ hours.



1.5 times the interquartile range is $1.5(8) = 12$ hours.
 A value less than $10 - 12 = -2$ hours is an outlier or a value greater than $18 + 12 = 30$ hours is an outlier.
 There are no outliers in this data set.

12. Ordered data: \$95, \$95, \$105, \$110, \$114, \$124, \$124, \$190, \$300

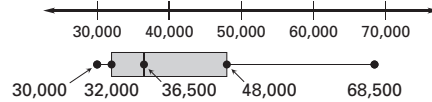
The median is \$114.
 The lower quartile is $\frac{95 + 105}{2} = \$100$.
 The upper quartile is $\frac{124 + 190}{2} = \$157$.
 The interquartile range is $157 - 100 = \$57$.



1.5 times the interquartile range is $1.5(57) = \$85.5$.
 A value less than $100 - 85.5 = \$14.5$ is an outlier and a value greater than $157 + 85.5 = \$242.5$ is an outlier.
 $300 > 242.5$, so \$300 is an outlier.

13. Ordered data: \$30,000, \$32,000, \$35,000, \$38,000, \$48,000, \$68,500

The median is $\frac{35,000 + 38,000}{2} = \$36,500$.
 The lower quartile is \$32,000.
 The upper quartile is \$48,000.
 The interquartile range is $48,000 - 32,000 = \$16,000$.



1.5 times the interquartile range is $1.5(16,000) = \$24,000$. A value less than $32,000 - 24,000 = \$8,000$ is an outlier. A value greater than $48,000 + 24,000 = \$72,000$ is an outlier.
 There are no outliers in this data set.

14. *Sample answer:* It is possible for two data sets to have the same mean, interquartile range, and range. The data sets *A* and *B* show this is true.

A: 1, 4, 4, 5, 7, 7, 8, 8, 10

B: 2, 3, 6, 6, 7, 7, 11

The mean of *A* is $\frac{54}{9} = 6$. The mean of *B* is $\frac{42}{7} = 6$.

The interquartile range of *A* is $8 - 4 = 4$.

The interquartile range of *B* is $7 - 3 = 4$.

The range of *A* is $10 - 1 = 9$.

The range of *B* is $11 - 2 = 9$.

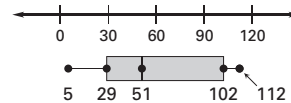
The data sets *A* and *B* have the same mean, interquartile range, and range.

15. Ordered data: 5, 11, 29, 37, 45, 51, 52, 79, 102, 105, 112

The median is 51 hours.

The lower quartile is 29 hours.

The upper quartile is 102 hours.

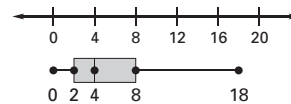


16. Ordered data: 0, 0, 1, 1, 2, 2, 2, 2, 4, 4, 5, 7, 8, 8, 8, 9, 9, 11, 18

The median is 4 runs.

The lower quartile is 2 runs.

The upper quartile is 8 runs.

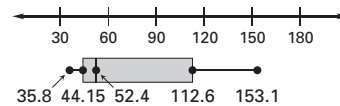


17. a. Ordered data: \$35.8, \$38.4, \$49.9, \$50.7, \$52.4, \$54.7, \$107.0, \$118.2, \$153.1

The median is \$52.4.

The lower quartile is $\frac{38.4 + 49.9}{2} = \44.15 .

The upper quartile is $\frac{107.0 + 118.2}{2} = \112.6 .

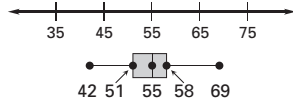


b. The interquartile range is $112.6 - 44.15 = \$68.45$. 1.5 times the interquartile range is $1.5(\$68.45) = \102.675 . A value that is less than $44.15 - 102.675 = -58.525$ or greater than $112.6 + 102.675 = \$215.275$ is an outlier. None of the states had retail sales in 2002 that can be considered outliers.

18. a. The median is 55 years.

The lower quartile is 51 years.

The upper quartile is 58 years.



b. The interquartile range is $58 - 51 = 7$ years. 1.5 times the interquartile range is $1.5(7) = 10.5$ years. A value that is less than $51 - 10.5 = 40.5$ or greater than $58 + 10.5 = 68.5$ is an outlier $69 > 68.5$, so 69 is an outlier. Ronald Reagan's age can be considered an outlier.

19. a. For Calisto, the interquartile range is $121.5 - 58.8$, or 63 kilometers. For Ganymede, the interquartile range is $93 - 50$, or 43 kilometers. So, Calisto has a greater variation in the middle 50% of the data. The range for Calisto, $980 - 38$, or 942 kilometers, is greater than the range for Ganymede, $588 - 16$, or 572 kilometers. So, Calisto has more variation in the diameter of its craters.

b. Chesapeake Bay's diameter of 90 kilometers is between the median and upper quartile of the diameters of craters on both Calisto and Ganymede. So, the diameter of Chesapeake Bay is larger than between 50% and 75% of craters on both Calisto and Ganymede.

c. Vredefort's diameter of 300 kilometers is between the upper quartile and the maximum value of the diameters of craters on both Calisto and Ganymede. So, Vredefort is larger than at least 75% of the craters on both Calisto and Ganymede.

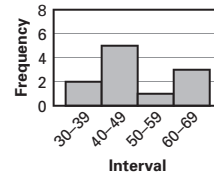
20. For the sopranos, 25% of the singers have a height greater than 66 inches. For the altos, between 25% and 50% of the singers have a height greater than 66 inches. For the tenors, about 75% of the singers have a height greater than 66 inches. For the bass, between 75% and 100% of the singers have a height greater than 66 inches. So, as voice pitch decreases, the height of the singers increases.

Quiz for the lessons "Interpret Stem-and-Leaf Plots and Histograms", and "Interpret Box-and-Whisker Plots"

- The total number of boys and girls interested in soccer is 122. The total number of boys and girls interested in lacrosse is 112. Since 122 is greater than 112, more ninth graders are interested in soccer.
- The total number of boys interested in lacrosse is 60. The total number of girls interested in lacrosse is 52. Since 60 is greater than 52, more boys are interested in lacrosse.

3.

Interval	Frequency
30-39	
40-49	
50-59	
60-69	



4. Stem | Leaves

1	28 8
2	22 8
3	36

Key: 2 | 2 = 2.2

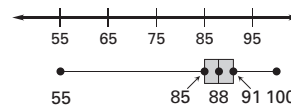
5. Ordered data: 55, 76, 79, 85, 85, 86, 88, 88, 88, 90, 90, 91, 92, 92, 100

The median is 88.

The lower quartile is 85.

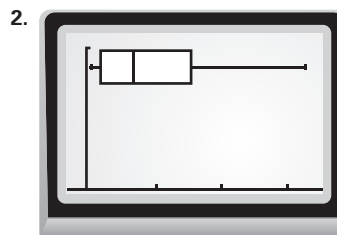
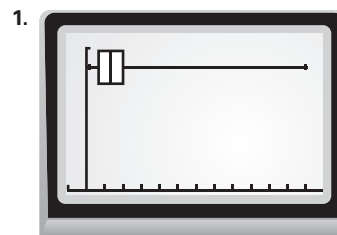
The upper quartile is 91.

The interquartile range is $91 - 85 = 6$.



1.5 times the interquartile range is $1.5(6) = 9$. A value that is less than $85 - 9 = 76$ or greater than $91 + 9 = 100$ is considered an outlier. The score of 55 is the only outlier in this data set.

Problem Solving Workshop



Extension for the lesson "Interpret Box-and-Whisker Plots"

- Since the data are close together with no outliers, it is appropriate to use a histogram. The mean is 15.44 and the standard deviation is about 3.49.

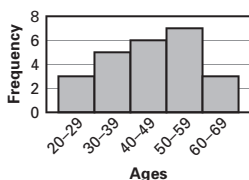
- Since three of the values appear to be much larger than the rest, it is appropriate to use a box-and-whisker plot. The median is 29 points and interquartile range is 17 points.
- Since the data are close together with no outliers, it is appropriate to use a histogram. The mean is 9.51 minutes and standard deviation is about 1.44 minutes.
- Since one of the values appears to be much smaller than the rest, it is appropriate to use a box-and-whisker plot. The median is 10.5 hours and the interquartile range is 1.75 hours.
- Since the data are close together with no outliers, it is appropriate to use a histogram. The mean is 25.7 cookies and the standard deviation is about 6.27 cookies.
- Since the data are close together with no outliers, it is appropriate to use a histogram. The mean is 35,890.75 people and the standard deviation is about 3161.98 people.

Mixed Review of Problem Solving for the lessons “Analyze Surveys and Samples,” “Use Measures of Central Tendency and Dispersion,” “Interpret Stem-and-Leaf Plots and Histograms,” and “Interpret Box-and-Whisker Plots”

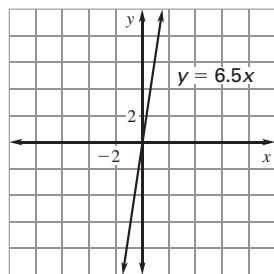
1. a.

Ages	Frequency
20–29	III
30–39	IIII
40–49	IIII I
50–59	IIII II
60–69	III

b.



2. a.



- b. The largemouth bass; *Sample answer:* The range and interquartile range for the largemouth bass data are both significantly greater than those for the blue gill.

- Jo sold 215 adult tickets.
 - Jo and Abe sold 782 student and adult tickets.
 - Jo sold 178 student tickets and Abe sold 201 student tickets. Since 201 is greater than 178, Abe sold more student tickets.

Chapter Review for the chapter “Data Analysis”

- An event that combines two or more events is a *compound event*.
- A possible result of an experiment is an *outcome*.
- Theoretical probability is based on knowing the likelihood of all possible outcomes of an event and compares the number of favorable outcomes to the number of possible outcomes. Experimental probability is based on trials of an experiment and compares the number of successes to the number of trials.

$$4. P(\text{black checker}) = \frac{\text{Number of black checkers}}{\text{Total number of checkers}} = \frac{15}{31}$$

$$5. P(\text{choosing } s) = \frac{\text{Number of tiles with } s}{\text{Total number of tiles}} = \frac{2}{9}$$

$$6. {}_7P_6 = \frac{7!}{(7-6)!} = \frac{7!}{1!} = 7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$$

$$7. {}_6P_2 = \frac{6!}{(6-2)!} = \frac{6!}{4!} = \frac{6 \cdot 5 \cdot 4!}{4!} = 30$$

$$8. {}_8P_5 = \frac{8!}{(8-5)!} = \frac{8!}{3!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3!}{3!} = 6720$$

$$9. {}_{13}P_{10} = \frac{13!}{(13-10)!} = \frac{13!}{3!} = \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3!}{3!} = 1,037,836,800$$

10. The number of possible outcomes is the number of permutations of 6 songs taken 4 at a time. This is ${}_6P_4$.

$${}_6P_4 = \frac{6!}{(6-4)!} = \frac{6!}{2!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2!}{2!} = 360$$

Only one of the permutations is the correct order, so there is one favorable outcome.

$$P(\text{first 4 songs}) = \frac{1}{360}$$

The probability of playing the first 4 songs you downloaded is $\frac{1}{360}$.

$$11. {}_7C_6 = \frac{7!}{(7-6)!6!} = \frac{7 \cdot 6!}{1! \cdot 6!} = 7$$

$$12. {}_6C_2 = \frac{6!}{(6-2)!2!} = \frac{6 \cdot 5 \cdot 4!}{4! \cdot (2 \cdot 1)} = 15$$

$$13. {}_8C_5 = \frac{8!}{(8-5)!5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{(3 \cdot 2 \cdot 1) \cdot 5!} = 56$$

$$14. {}_{13}C_{10} = \frac{13!}{(13-10)!10!} = \frac{13 \cdot 12 \cdot 11 \cdot 10!}{(3 \cdot 2 \cdot 1) \cdot 10!} = 286$$

15. The order in which you choose your friends is not important. So, to find the number of combinations of 9 friends taken 4 at a time, find ${}_9C_4$.

$${}_9C_4 = \frac{9!}{(9-4)!4!} = \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5!}{5! \cdot (4 \cdot 3 \cdot 2 \cdot 1)} = 126$$

There are 126 ways you can choose 4 of your friends.

16. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
 $= 0.32 + 0.48 - 0.12$
 $= 0.68$

17. $P(\bar{A}) = 1 - P(A) = 1 - 0.32 = 0.68$

18. $P(\bar{B}) = 1 - P(B) = 1 - 0.48 = 0.52$

19. a. $P(\text{red and green}) = P(\text{red}) \cdot P(\text{green})$
 $= \frac{5}{16} \cdot \frac{8}{16} = \frac{5}{32}$

b. $P(\text{red and green}) = P(\text{red}) \cdot P(\text{green} | \text{red})$
 $= \frac{5}{16} \cdot \frac{8}{15} = \frac{1}{6}$

20. a. $P(\text{blue and red}) = P(\text{blue}) \cdot P(\text{red})$
 $= \frac{3}{16} \cdot \frac{5}{16} = \frac{15}{256}$

b. $P(\text{blue and red}) = P(\text{blue}) \cdot P(\text{red} | \text{blue})$
 $= \frac{3}{16} \cdot \frac{5}{15} = \frac{1}{16}$

21. a. $P(\text{green and green}) = P(\text{green}) \cdot P(\text{green})$
 $= \frac{8}{16} \cdot \frac{8}{16} = \frac{1}{4}$

b. $P(\text{green and green}) = P(\text{green}) \cdot P(\text{green} | \text{green})$
 $= \frac{8}{16} \cdot \frac{7}{15} = \frac{7}{30}$

Chapter Test for the chapter "Data Analysis"

1. a. $P(4) = \frac{1}{6}$

b. Odds in favor of a 4 = $\frac{1}{5}$ or 1:5

2. a. $P(\text{even}) = \frac{3}{6} = \frac{1}{2}$

b. Odds in favor of even = $\frac{3}{3} = \frac{1}{1}$ or 1:1

3. a. $P(\text{less than 5}) = \frac{4}{6} = \frac{2}{3}$

b. Odds in favor of less than 5 = $\frac{4}{2} = \frac{2}{1}$ or 2:1

4. a. $P(\text{multiple of 3}) = \frac{2}{6} = \frac{1}{3}$

b. Odds in favor of a multiple of 3 = $\frac{2}{4} = \frac{1}{2}$ or 1:2

5. ${}_7P_2 = \frac{7!}{(7-2)!} = \frac{7!}{5!} = \frac{7 \cdot 6 \cdot 5!}{5!} = 42$

6. ${}_8P_3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{5!} = 336$

7. ${}_6C_3 = \frac{6!}{(6-3)!3!} = \frac{6 \cdot 5 \cdot 4 \cdot 3!}{(3 \cdot 2 \cdot 1) \cdot 3!} = 20$

8. ${}_{12}C_7 = \frac{12!}{(12-7)!7!} = \frac{12!}{5!7!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7!}{(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)7!}$
 $= 792$

9. The order that the swimmers finish is important so permutations must be used. The number of ways the 8 swimmers can finish is the number of permutations of 8 swimmers taking 3 at a time, or ${}_8P_3$.

$${}_8P_3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{5!} = 336$$

There are 336 ways the swimmers can finish in first, second, and third place.

10. The order of the sides does not matter so combinations must be used. The number of ways to choose 2 side dishes is the number combination of 7 sides taken 2 at a time, or ${}_7C_2$.

$${}_7C_2 = \frac{7!}{(7-2)!2!} = \frac{7 \cdot 6 \cdot 5!}{5! \cdot (2 \cdot 1)} = 21$$

There are 21 ways to choose 2 sides out of 7 different side dishes.

11. $P(\text{less than 10 or odd}) = P(\text{less than 10} + P(\text{odd}) - P(\text{less than 10 and odd}))$

$$= \frac{9}{12} + \frac{6}{12} - \frac{5}{12} = \frac{10}{12} = \frac{5}{6}$$

12. $P(\text{greater than 3 then odd}) = P(\text{greater than 3}) \cdot P(\text{odd})$

$$= \frac{9}{12} \cdot \frac{6}{12} = \frac{54}{144} = \frac{3}{8}$$

13. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
 $= 0.3 + 0.6 - 0.1$
 $= 0.8$

14. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
 $80\% = 35 + P(B) - 20\%$
 $65\% = P(B)$

15. $P(\bar{A}) = 1 - P(A)$

$$\frac{2}{5} = 1 - P(A)$$

$$P(A) = 1 - \frac{2}{5} = \frac{3}{5}$$

16. $P(A \text{ and } B) = P(A) \cdot P(B) = (0.15)(0.6) = 0.09$

17. $P(A \text{ and } B) = P(A) \cdot P(B | A)$
 $25\% = 60\% \cdot P(B | A)$
 $41.7\% \approx P(B | A)$

18. $P(A \text{ and } B) = P(A) \cdot P(B | A)$
 $0.36 = P(A) \cdot 0.4$
 $0.9 = P(A)$

19. $P(\text{female not in activity}) = \frac{\text{Females not in activity}}{\text{Total number of students}}$
 $= \frac{325}{1800} \approx 0.18$

The probability that a randomly-selected student is a female who is not involved in an activity is about 0.18.

Extra Practice for the chapter "Data Analysis"

- The population is parents or guardians of high school students. The sampling method is systematic sampling.
- Sample answer:* The sampling method is not likely to result in a biased sample. The rule used to choose the individuals is not biased, so the sample will most likely be representative of the population.
- Sample answer:* The question is potentially biased because it suggests the sound system needs updating.

4. $\bar{x} = \frac{101 + 88 \dots 78}{10} = \frac{890}{10} = 89$

The median is $\frac{88 + 88}{2} = 88$.

The mode is 88.

The range is $108 - 69 = 39$.

The mean is 89, so the mean absolute deviation is:

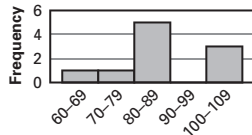
$$\frac{|101 - 89| + |88 - 89| + \dots + |78 - 89|}{10} = 8.8$$

5. fries; salad; fries

6. $\frac{68 + 43}{68 + 76 + 43} = \frac{111}{187} = 0.59358\dots \approx 59\%$

7.

Interval	Frequency
69-78	II
79-88	IIII
89-98	
99-108	III



Stem	Leaves
6	9
7	8
8	0 8 8 8 8
9	
10	1 2 8

Key: 6|9 = 69

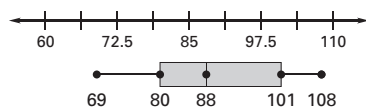
Key: $\frac{7}{8} = 78$

8. *Ordered data:* 69, 78, 80, 88, 88, 88, 88, 101, 102, 108

The median is 88.

The lower quartile is 80.

The upper quartile is 101.



The interquartile range is $101 - 80 = 21$.

A number that is less than $80 - 1.5(21) = 48.5$ or greater than $101 + 1.5(21) = 132.5$ is an outlier. There are no outliers for this data set.

- The box-and-whisker plot shows 13.5 as the median; 13,500 ft
- The interquartile range is the difference between the 3rd and 1st quartiles: $14.4 - 12.7 = 1.7$; 1,700 ft

HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT
HOUGHTON MIFFLIN HARCOURT HOUGHTON MIFFLIN HARCOURT