14-3 Measures of Variability

OBJECTIVES

- Find the interquartile range, the semi-interquartile range, mean deviation, and standard deviation of a set of data.
- Organize and compare data using box-and-whisker plots.

EDUCATION Are you planning to attend college? If so, do you know which school you are going to attend? There are several factors influencing students' decisions concerning which college to attend. Two of those factors may be the cost of tuition and the size of the school. The table lists some of the largest colleges with their total enrollment and cost for in-state tuition and fees.

College Enrollment and Tuition

College	Enrollment, 1997-1998	Tuition and Fees ($), 1997-1998
University of Texas	47,476	2866
The Ohio State University	45,462	3687
Penn State University	37,718	5832
University of Georgia	29,693	2838
Florida State University	28,285	1988
University of Southern California	27,874	20,480
Virginia Tech	24,481	4147
North Carolina State University	24,141	2232
Texas Tech University	24,075	2414
University of South Carolina	22,836	3534
University of Nebraska	22,393	2769
Colorado State University	21,970	2933
University of Illinois	21,645	4364
Auburn University (AL)	21,498	2610
University of Kentucky	20,925	2736
Kansas State University	20,325	2467
University of Oklahoma	19,886	2311
Cornell University (NY)	18,001	21,914
University of Alaska	17,090	2294

Source: College Entrance Examination Board

interNET CONNECTION

Data Update For the latest information about college enrollment and tuition, visit **www.amc.glencoe.com**

You will solve problems related to this in Examples 1-4.

Measures of central tendency, such as the mean, median, and mode, are statistics that describe certain important characteristics of data. However, they do not indicate anything about the variability of the data. For example, 50 is the mean of both {0, 50, 100} and {40, 50, 60}. The variability is much greater in the first set of data than in the second, since $100 - 0$ is much greater than $60 - 40$.

One **measure of variability** is the *range*. Use the information in the table above to find the range of enrollment.

$$\underbrace{47{,}476}_{\text{University of Texas}} - \underbrace{17{,}090}_{\text{University of Alaska}} = 30{,}386.$$

The range of enrollment is 30,386 students.

If the median is a member of the set of data, that item of data is excluded when calculating the first and third quartile points.

If the data have been arranged in order and the median is found, the set of data is divided into two groups. Then if the median of each group is found, the data is divided into four groups. Each of these groups is called a **quartile.** There are three quartile points, Q_1, Q_2, and Q_3, that denote the breaks in the data for each quartile. The median is the second quartile point Q_2. The medians of the two groups defined by the median are the first quartile point Q_1 and the third quartile point Q_3.

One fourth of the data is less than the first quartile point Q_1, and three fourths of the data is less than the third quartile point Q_3. The difference between the first quartile point and third quartile point is called the **interquartile range.** When the interquartile range is divided by 2, the quotient is called the **semi-interquartile range.**

Semi-Interquartile Range

If a set of data has first quartile point Q_1 and third quartile point Q_3, the semi-interquartile range Q_R can be found as follows.

$$Q_R = \frac{Q_3 - Q_1}{2}$$

Example 1 **EDUCATION** **Refer to the application at the beginning of the lesson.**

Real World Application

a. Find the interquartile range of the college enrollments and state what it represents.

b. Find the semi-interquartile range of the college enrollments.

a. First, order the data from least to greatest, and identify Q_1, Q_2, and Q_3.

17,090 18,001 19,886 20,325 <u>20,925</u> (Q_1) 21,498 21,645 21,970 22,393 <u>22,836</u> (Q_2)

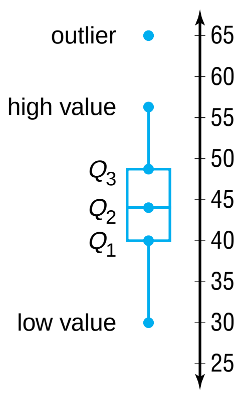
24,075 24,141 24,481 27,874 <u>28,285</u> (Q_3) 29,693 37,718 45,462 47,476

The interquartile range is 28,285 − 20,925 or 7360. This means that the middle half of the student enrollments are between 28,285 and 20,925 and are within 7360 of each other.

b. The semi-interquartile range is $\frac{7360}{2}$ or 3680. The halfway point between Q_1 and Q_3 can be found by adding the semi-interquartile range to Q_1. That is, 3680 + 20,925 or 24,605. Since 24,605 > Q_2, this indicates the data is more clustered between Q_1 and Q_2 than between Q_2 and Q_3.

Graphing Calculator Tip

Enter the data into **L1** and use the **SortA(** command to reorder the list from least to greatest.



Box-and-whisker plots are used to summarize data and to illustrate the variability of the data. These plots graphically display the median, quartiles, interquartile range, and extreme values in a set of data. They can be drawn vertically, as shown at the right, or horizontally. A box-and-whisker plot consists of a rectangular box with the ends, or **hinges,** located at the first and third quartiles. The segments extending from the ends of the box are called **whiskers.** The whiskers stop at the extreme values of the set, unless the set contains **outliers.** Outliers are extreme values that are more than 1.5 times the interquartile range beyond the upper or lower quartiles. Outliers are represented by single points. If an outlier exists, each whisker is extended to the last value of the data that is not an outlier.

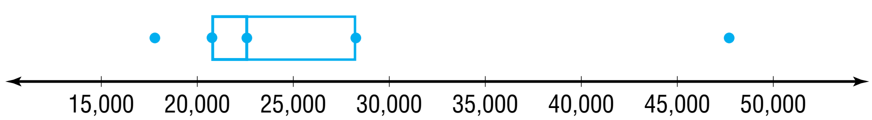
The dimensions of the box-and-whisker plot can help you characterize the data. Each whisker and each small box contains 25% of the data. If the whisker or box is short, the data are concentrated over a narrower range of values. The longer the whisker or box, the larger the range of the data in that quartile. Thus, the box-and-whisker is a pictorial representation of the variability of the data.

Example 2 Real World Application

EDUCATION **Refer to the application at the beginning of the lesson. Draw a box-and-whisker plot for the enrollments.**

In Example 1, you found that Q_1 is 20,925, Q_2 is 22,836, and Q_3 is 28,285. The extreme values are the least value 17,090 and the greatest value 47,476.

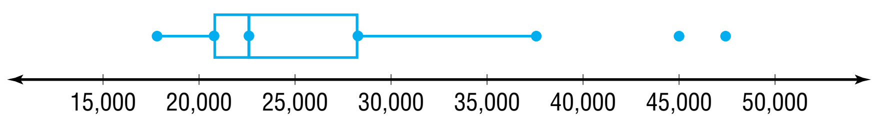
Draw a number line and plot the quartiles, the median, and the extreme values. Draw a box to show the interquartile range. Draw a segment through the median to divide the box into two smaller boxes.



Before drawing the whiskers, determine if there are any outliers. From Example 1, we know that the interquartile range is 7360. An outlier is any value that lies more than 1.5(7360) or 11,040 units below Q_1 or above Q_3.

$$Q_1 - 1.5(7360) = 20{,}925 - 11{,}040 = 9885 \qquad Q_3 + 1.5(7360) = 28{,}285 + 11{,}040 = 39{,}325$$

The lower extreme 17,090 is within the limits. However, 47,476 and 45,462 are not within the limits. They are outliers. Graph these points on the plot. Then draw the left whisker from 17,090 to 20,925 and the right whisker from 28,285 to the greatest value that is not an outlier, 37,718.



The box-and-whisker plot shows that the two lower quartiles of data are fairly concentrated. However, the upper quartile of data is more diverse.

Another measure of variability can be found by examining deviation from the mean, symbolized by $X_i - \overline{X}$. The sum of the deviations from the mean is zero. That is, $\sum_{i=1}^{n} (X_i - \overline{X}) = 0$. For example, the mean of the data set {14, 16, 17, 20, 33} is 20. The sum of the deviations from the mean is shown in the table.

X_i	$\overline{X}$	$X_i - \overline{X}$
14	20	−6
16	20	−4
17	20	−3
20	20	0
33	20	13
$\sum_{i=1}^{5} (X_i - \overline{X})$		0

To indicate how far individual items vary from the mean, we use the absolute values of the deviation. The arithmetic mean of the absolute values of the deviations from the mean of a set of data is called the **mean deviation,** symbolized by *MD*.

Mean Deviation

If a set of data has n values given by X_i, such that $1 \le i \le n$, with arithmetic mean $\overline{X}$, then the mean deviation MD can be found as follows.

$$MD = \frac{1}{n} \sum_{i=1}^{n} |X_i - \overline{X}|$$

In sigma notation for statistical data, i is always an integer and not the imaginary unit.

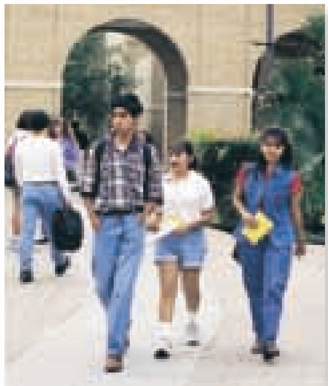
Example 3

Real World Application

EDUCATION **Refer to the application at the beginning of the lesson. Find the mean deviation of the enrollments.**

There are 19 college enrollments listed, and the mean is $\frac{1}{19}\sum_{i=1}^{19} X_i$ or about 26,093.37.

Method 1: Sigma notation



$$MD \approx \frac{1}{19}\sum_{i=1}^{19} |X_i - 26{,}093.37|$$

$$MD \approx \frac{1}{19}\Big(|47{,}476 - 26{,}093.37| + |45{,}462 - 26{,}093.37| + \cdots + |17{,}090 - 26{,}093.37|\Big)$$

$$MD \approx \frac{1}{19}\Big(|21{,}382.63| + |19{,}368.63| + \cdots + |-9003.37|\Big)$$

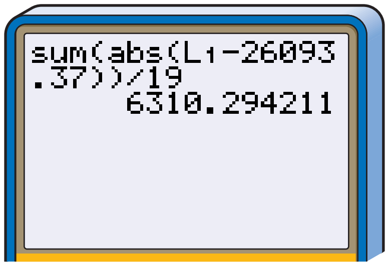
$$MD \approx 6310.29$$

The mean deviation of the enrollments is about 6310.29. This means that the enrollments are an average of about 6310.29 above or below the mean enrollment of 26,093.37.

Method 2: Graphing Calculator

Enter the data for the enrollments into **L1**. At the home screen, enter the following formula.

sum(abs(L1 − 26093.37))/19



The calculator determines the difference between the scores and the mean, takes the absolute value, adds the absolute values of the differences, and divides by 19. This verifies the calculation in Method 1.

Graphing Calculator Tip

The **sum(** command is located in the **MATH** section of the **LIST** menu. The **abs(** command is in the **NUM** section after pressing **MATH**.

A measure of variability that is often associated with the arithmetic mean is the **standard deviation.** Like the mean deviation, the standard deviation is a measure of the average amount by which individual items of data deviate from the arithmetic mean of all the data. Each individual deviation can be found by subtracting the arithmetic mean from each individual value, $X_i - \overline{X}$. Some of these differences will be negative, but if they are squared, the results are positive. The standard deviation is the square root of the mean of the squares of the deviation from the arithmetic mean.

Standard Deviation	If a set of data has n values, given by X_i such that $1 \leq i \leq n$, with arithmetic mean $\overline{X}$, the standard deviation σ can be found as follows. $$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

σ is the lowercase Greek letter sigma.

The standard deviation is the most important and widely used measure of variability. Another statistic used to describe the spread of data about the mean is **variance.** The variance, denoted σ^2, is the mean of the squares of the deviations from $\overline{X}$. The standard deviation is the positive square root of the variance.

Example 4 Real World Application

EDUCATION **Refer to the application at the beginning of the lesson. Find the standard deviation of the enrollments.**

Method 1: Standard Deviation Formula

There are 19 college enrollments listed, and the mean is about 26,093.37.

$$\sigma \approx \sqrt{\frac{1}{19}\sum_{i=1}^{19}(X_i - 26{,}093.37)^2}$$

$$\sigma \approx \sqrt{\frac{1}{19}(47{,}476 - 26{,}093.37)^2 + (45{,}462 - 26{,}093.37)^2 + \cdots + (17{,}090 - 26{,}093.37)^2}$$

$$\sigma \approx \sqrt{\frac{1}{19}(21{,}382.63)^2 + (19{,}368.63)^2 + \cdots + (-9003.37)^2}$$

$$\sigma \approx 8354.59$$

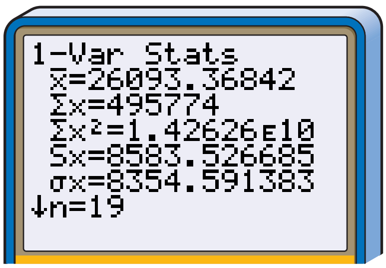
The standard deviation is about 8354.59. Since the mean of the enrollments is about 26,093.37 and the standard deviation is about 8354.59, the data have a great amount of variability.

Method 2: Graphing Calculator

Enter the data in **L1**. Use the **CALC** menu after pressing STAT to find the 1-variable statistics.

The standard deviation, indicated by **σx**, is the fifth statistic listed.

The mean ($\bar{x}$) is 26,093.36842 and the standard deviation is 8354.5913383, which agree with the calculations using the formulas.



When studying the standard deviation of a set of data, it is important to consider the mean. For example, compare a standard deviation of 5 with a mean of 10 to a standard deviation of 5 with a mean of 1000. The latter indicates very little variation, while the former indicates a great deal of variation since 5 is 50% of 10 while 5 is only 0.5% of 1000.

The standard deviation of a frequency distribution is the square root of the mean of the squares of the deviations of the class marks from the mean of the frequency data, weighted by the frequency of each interval.

Standard Deviation of the Data in a Frequency Distribution

If $X_1, X_2, \ldots, X_k$ are the class marks in a frequency distribution with k classes, and $f_1, f_2, \ldots, f_k$ are the corresponding frequencies, then the standard deviation σ of the data in the frequency distribution is found as follows.

$$\sqrt{\frac{\sum_{i=1}^{k}(X_i - \overline{X})^2 \cdot f_i}{\sum_{i=1}^{k} f_i}}$$

The standard deviation of a frequency distribution is an approximate number.

Example 5 **ECONOMICS** **Use the frequency distribution data below to find the arithmetic mean and the standard deviation of the price-earnings ratios of 100 manufacturing stocks.**

Real World Application

Method 1: Using Formulas

Class Limits	Class Marks (X)	f	$f \cdot X$	$(X - \overline{X})$	$(X - \overline{X})^2$	$(X - \overline{X})^2 \cdot f$
−0.5–4.5	2.0	5	10	−8	64	320
4.5–9.5	7.0	54	378	−3	9	486
9.5–14.5	12.0	25	300	2	4	100
14.5–19.5	17.0	13	221	7	49	637
19.5–24.5	22.0	0	0	12	144	0
24.5–29.5	27.0	1	27	17	289	289
29.5–34.5	32.0	2	64	22	484	968
		100	1000			2800

The mean $\overline{X}$ is $\frac{1000}{100}$ or 10.

The standard deviation σ is $\sqrt{\frac{2800}{100}}$ or approximately 5.29.

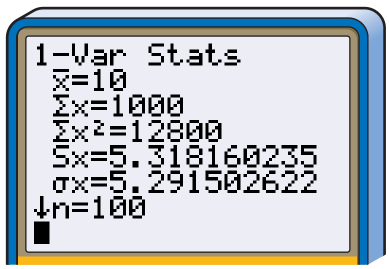
Since the mean number of price-earnings ratios is 10 and the standard deviation is 5.29, this indicates a great amount of variability in the data.

Method 2: Graphing Calculator

Enter the class marks in the **L1** list and the frequency in the **L2** list.

Use the **CALC** menu after pressing STAT to find the 1-variable statistics. Then type **L1**, **L2** and press ENTER.

The calculator confirms the standard deviation is about 5.29.

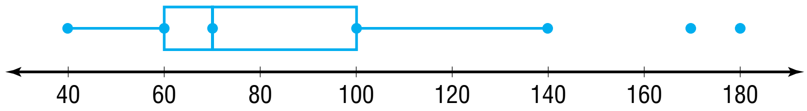


Check for Understanding

Communicating Mathematics

Read and study the lesson to answer each question.

1. **Describe** the data shown in the box-and-whisker plot below. Include the quartile points, interquartile range, semi-interquartile range, and any outliers.



2. **Explain** how to find the variance of a set of data if you know the standard deviation.
3. **Compare and contrast** mean deviation and standard deviation.
4. *Math Journal* **Draw** a box-and-whisker plot for data you found in a newspaper or magazine. What conclusions can you derive from the plot?

Guided Practice

5. Find the interquartile range and the semi-interquartile range of {17, 28, 44, 37, 28, 42, 21, 41, 35, 25}. Then draw a box-and-whisker plot.
6. Find the mean deviation and the standard deviation of {\$4.45, \$5.50, \$5.50, \$6.30, \$7.80, \$11.00, \$12.20, \$17.20}
7. Find the arithmetic mean and the standard deviation of the frequency distribution at the right.

Class Limits	Frequency
0–10,000	15
10,000–20,000	30
20,000–30,000	50
30,000–40,000	60
40,000–50,000	30
50,000–60,000	15

8. **Meteorology** The following table gives the normal maximum daily temperature for Los Angeles and Las Vegas.

Normal Maximum Daily Temperatures

	January	February	March	April	May	June
Los Angeles	65.7	65.9	65.5	67.4	69.0	71.9
Las Vegas	57.3	63.3	68.8	77.5	87.8	100.3

	July	August	September	October	November	December
Los Angeles	75.3	76.6	76.6	74.4	70.3	65.9
Las Vegas	105.9	103.2	94.7	82.1	67.4	57.5

Source: National Oceanic and Atmosphere Administration

a. Find the mean, median, and standard deviation for the temperatures in Los Angeles.
b. What are the mean, median, and standard deviation for the temperatures in Las Vegas?
c. Draw a box-and-whisker plot for the temperatures for each city.
d. Which city has a smaller variability in temperature?
e. What might cause one city to have a greater variability in temperature than another?

EXERCISES

Practice

Find the interquartile range and the semi-interquartile range of each set of data. Then draw a box-and-whisker plot.

9. {30, 28, 24, 24, 22, 22, 21, 17, 16, 15}

10. {7, 14, 18, 72, 13, 15, 19, 8, 17, 28, 11, 15, 24}

11. {15.1, 9.0, 8.5, 5.8, 6.2, 8.5, 10.5, 11.5, 8.8, 7.6}

12. Use a graphing calculator to draw a box-and-whisker plot for {7, 1, 11, 5, 4, 8, 12, 15, 9, 6, 5, 9}?

Find the mean deviation and the standard deviation of each set of data.

13. {200, 476, 721, 579, 152, 158}

14. {5.7, 5.7, 5.6, 5.5, 5.3, 4.9, 4.4, 4.0, 4.0, 3.8}

15. {369, 398, 381, 392, 406, 413, 376, 454, 420, 385, 402, 446}

16. Find the variance of {34, 55, 91, 13, 22}.

Find the arithmetic mean and the standard deviation of each frequency distribution.

17.

Class Limits	Frequency
1–5	2
5–9	8
9–13	15
13–17	6
17–21	38
21–25	31
25–29	13
29–33	7

18.

Class Limits	Frequency
53–61	3
61–69	7
69–77	11
77–85	38
85–93	19
93–101	12

19.

Class Limits	Frequency
70–90	2
90–110	11
110–130	39
130–150	17
150–170	9
170–190	7

Applications and Problem Solving

Real World Application

20. Geography There are seven navigable rivers that feed into the Ohio River. The lengths of these rivers are given at the right.

a. Find the median of the lengths.

b. Name the first quartile point and the third quartile point.

c. Find the interquartile range.

d. What is the semi-interquartile range?

e. Are there any outliers? If so, name them.

f. Make a box-and-whisker plot of the lengths of the rivers.

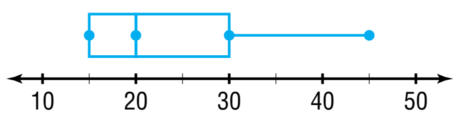
g. Use the box-and-whisker plot to discuss the variability of the data.

Length of Rivers Feeding into the Ohio River

River	Length
Monongahela	129 miles
Allegheny	325 miles
Kanawha	97 miles
Kentucky	259 miles
Green	360 miles
Cumberland	694 miles
Tennessee	169 miles

Source: *The Universal Almanac*

21. Critical Thinking Write a set of numerical data that could be represented by the box-and-whisker plot at the right.



22. **Sports** During a recent season, 7684 teams played 19 NCAA women's sports. The breakdown of these teams is given below.

Sport	Teams	Sport	Teams	Sport	Teams
Basketball	966	Lacrosse	182	Swimming	432
Cross Country	838	Rowing	97	Tennis	859
Fencing	42	Skiing	40	Track, Indoor	528
Field Hockey	228	Soccer	691	Track, Outdoor	644
Golf	282	Softball	770	Volleyball	923
Gymnastics	91	Squash	26	Water Polo	23
Ice Hockey	22				

Source: The National Collegiate Athletic Association

a. What is the median of the number of women's teams playing a sport?
b. Find the first quartile point and the third quartile point.
c. What is the interquartile range and semi-interquartile range?
d. Are there any outliers? If so, name them.
e. Make a box-and-whisker of the number of women's teams playing a sport.
f. What is the mean of the number of women's teams playing a sport?
g. Find the mean deviation of the data.
h. Find the variance of the data.
i. What is the standard deviation of the data?
j. Discuss the variability of the data.

23. **Education** Refer to the data on the college tuition and fees in the application at the beginning of the lesson.
a. What are the quartile points of the data?
b. Find the interquartile range.
c. Name any outliers.
d. Make a box-and-whisker plot of the data.
e. What is the mean deviation of the data?
f. Find the standard deviation of the data.
g. Discuss the variability of the data.

24. **Government** The number of times the first 42 presidents vetoed bills are listed below.

2, 0, 0, 7, 1, 0, 12, 1, 0, 10, 3, 0, 0, 9, 7, 6, 29, 93, 13, 0, 12, 414, 44, 170, 42, 82, 39, 44, 6, 50, 37, 635, 250, 181, 21, 30, 43, 66, 31, 78, 44, 25

a. Make a box-and-whisker plot of the number of vetoes.
b. Find the mean deviation of the data.
c. What is the variance of the data?
d. What is the standard deviation of the data?
e. Describe the variability of the data.

25. **Entertainment** The frequency distribution shows the average audience rating for the top fifty network television shows for one season.

Audience Rating	8–10	10–12	12–14	14–16	16–18	18–20	20–22
Frequency	26	12	6	2	2	0	2

Source: Nielsen Media Research

a. Find the arithmetic mean of the audience ratings.
b. What is the standard deviation of the audience ratings?

26. **Critical Thinking** Is it possible for the variance to be less than the standard deviation for a set of data? If so, explain when this will occur. When would the variance be equal to the standard deviation for a set of data?

27. **Research** Find the number of students attending each school in your county. Make a box-and-whisker plot of the data. Determine various measures of variability and discuss the variability of the data.

Mixed Review

28. Consider the data represented by the stem-and-leaf plot at the right. *(Lesson 14-2)*

stem	leaf
4	4 4 9
5	4 5
6	2 2 4 5 9
7	1 4 5 6 7 8 9
8	0 2 4 5 6 7 8 9 9 9
9	0 2 3 3 5 6 8 9

5|4 = 5.4

a. What is the mean of the data?

b. Find the median of the data.

c. What is the mode of the data?

29. **Fund-Raising** Twelve students are selling programs at the Grove City High School to raise money for the athletic department. The numbers of programs sold by each student are listed below. *(Lesson 14-1)*

51, 27, 55, 54, 68, 60, 39, 46, 46, 53, 57, 23

a. Find the range of the number of programs sold.

b. Determine an appropriate class interval.

c. What are the class limits?

d. Construct a frequency distribution of the data.

e. Draw a histogram of the data.

30. **Food Service** Suppose nine salad toppings are placed on a circular, revolving tray. How many ways can the salad items be arranged? *(Lesson 13-2)*

31. Find the first three iterates of the function $f(x) = 0.5x - 1$ using $x_0 = 8$. *(Lesson 12-8)*

32. **SAT/ACT Practice** A carpenter divides a board that is 7 feet 9 inches long into three equal parts. What is the length of each part?

A 2 ft $6\frac{1}{3}$ in. **B** 2 ft $8\frac{1}{3}$ in. **C** 2 ft 7 in.

D 2 ft 8 in. **E** 2 ft 9 in.

MID-CHAPTER QUIZ

The scores for an exam given in physics class are given below.

82, 77, 84, 98, 93, 71, 76, 64, 89, 95, 78, 89, 65, 88, 54, 96, 87, 92, 80, 85, 93, 89, 55, 62, 79, 90, 86, 75, 99, 62

1. What is an appropriate class interval for the test scores? (Lesson 14-1)

2. Construct a frequency distribution of the test scores. (Lesson 14-1)

3. Draw a histogram of the test scores. (Lesson 14-1)

4. Make a stem-and-leaf plot of the test scores. (Lesson 14-2)

5. What is the mean of the test scores? (Lesson 14-2)

6. Find the median of the test scores. (Lesson 14-2)

7. Find the mode of the test scores. (Lesson 14-2)

8. Make a box-and-whisker plot of the test scores. (Lesson 14-3)

9. What is the mean deviation of the test scores? (Lesson 14-3)

10. Discuss the variability of the data. (Lesson 14-3)

Extra Practice See p. A54.

CONTENTS