2.5 Measures of Position

What You SHOULD LEARN

- How to find the first, second, and third quartiles of a data set
- How to find the interquartile range of a data set
- How to represent a data set graphically using a box-and-whisker plot
- How to interpret other fractiles such as percentiles
- How to find and interpret the standard score (z-score)

Quartiles ▸ Percentiles and Other Fractiles ▸ The Standard Score

▸ Quartiles

In this section, you will learn how to use fractiles to specify the position of a data entry within a data set. **Fractiles** are numbers that partition, or divide, an ordered data set into equal parts. For instance, the median is a fractile because it divides an ordered data set into two equal parts.

DEFINITION

The three **quartiles,** Q_1, Q_2, and Q_3, approximately divide an ordered data set into four equal parts. About one quarter of the data fall on or below the **first quartile** Q_1. About one half of the data fall on or below the **second quartile** Q_2 (the second quartile is the same as the median of the data set). About three quarters of the data fall on or below the **third quartile** Q_3.

EXAMPLE 1

Finding the Quartiles of a Data Set

The test scores of 15 employees enrolled in a CPR training course are listed. Find the first, second, and third quartiles of the test scores.

13 9 18 15 14 21 7 10 11 20 5 18 37 16 17

Solution

First, order the data set and find the median Q_2. Once you find Q_2, divide the data set into two halves. The first and third quartiles are the medians of the lower and upper halves of the data set.

Lower half: 5 7 9 10 11 13 14
15
Upper half: 16 17 18 18 20 21 37

Q_1 = 10, Q_2 = 15, Q_3 = 18

Interpretation About one fourth of the employees scored 10 or less; about one half scored 15 or less; and about three fourths scored 18 or less.

▸ Try It Yourself 1

Find the first, second, and third quartiles for the number of touchdowns scored by all Division 1A football teams using the data set listed in the Chapter Opener on page 39.

a. *Order* the data set.
b. Find the median Q_2.
c. Find the first and third quartiles Q_1 and Q_3.

Answer: Page A36

EXAMPLE 2

Using Technology to Find Quartiles

The tuition costs (in thousands of dollars) for 25 liberal arts colleges are listed. Use a calculator or a computer to find the first, second, and third quartiles.

23 25 30 23 20 22 21 15 25 24 30 25 30
20 23 29 20 19 22 23 29 23 28 22 28

Solution MINITAB, Excel, and the TI-83/84 each have features that automatically calculate quartiles. Try using this technology to find the first, second, and third quartiles of the tuition data. From the displays, you can see that $Q_1 = 21.5$, $Q_2 = 23$, and $Q_3 = 28$.

Study Tip

There are several ways to find the quartiles of a data set. Regardless of how you find the quartiles, the results are rarely off by more than one data entry. For instance, in Example 2, the first quartile, as determined by Excel, is 22 instead of 21.5.

MINITAB

Descriptive Statistics

Variable	N	Mean	SE Mean	TrMean	StDev
Tuition	25	23.960	0.788	24.087	3.942

Variable	Minimum	Q1	Median	Q3	Maximum
Tuition	15.000	21.500	23.000	28.000	30.000

EXCEL

	A	B	C	D
1	23			
2	25		Quartile(A1:A25,1)	
3	30		22	
4	23			
5	20		Quartile(A1:A25,2)	
6	22		23	
7	21			
8	15		Quartile(A1:A25,3)	
9	25		28	
10	24			
11	30			
12	25			
13	30			
14	20			
15	23			
16	29			
17	20			
18	19			
19	22			
20	23			
21	29			
22	23			
23	28			
24	22			
25	28			

TI-83/84

1-Var Stats
↑n=25
minX=15
Q_1=21.5
Med=23
Q_3=28
maxX=30

Interpretation About one quarter of these colleges charge tuition of $21,500 or less; one half charge $23,000 or less; and about three quarters charge $28,000 or less.

▸ Try It Yourself 2

The tuition costs (in thousands of dollars) for 25 universities are listed. Use a calculator or a computer to find the first, second, and third quartiles.

20 26 28 25 31 14 23 15 12 26 29 24 31
19 31 17 15 17 20 31 32 16 21 22 28

a. *Enter* the data.
b. *Calculate* the first, second, and third quartiles.
c. What can you conclude?.

Answer: Page A36

After finding the quartiles of a data set, you can find the interquartile range.

DEFINITION

The **interquartile range (IQR)** of a data set is the difference between the third and first quartiles.

Interquartile range (IQR) $= Q_3 - Q_1$

EXAMPLE 3

Finding the Interquartile Range

Find the interquartile range of the 15 test scores given in Example 1. What can you conclude from the result?

Solution From Example 1, you know that $Q_1 = 10$ and $Q_3 = 18$. So, the interquartile range is

$$\text{IQR} = Q_3 - Q_1 = 18 - 10 = 8.$$

Interpretation The test scores in the middle portion of the data set vary by at most 8 points.

Try It Yourself 3

Find the interquartile range for the number of touchdowns scored by all Division 1A football teams listed in the Chapter Opener on page 39.

a. Find the first and third quartiles, Q_1 and Q_3.
b. *Subtract* Q_1 from Q_3.
c. *Interpret* the result in the context of the data.

Answer: Page A36

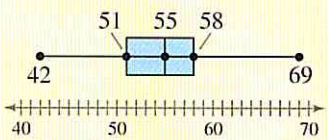
The IQR is a measure of variation that gives you an idea of how much the middle 50% of the data varies. It can also be used to identify outliers. Any data value that lies more than 1.5 IQRs to the left of Q_1 or to the right of Q_3 is an outlier. For instance, the IQR in Example 1 is $18 - 10 = 8$. So, 1.5 IQRs to the right of Q_3 is $Q_3 + 1.5(8) = 18 + 12 = 30$. Because $37 > 30$, 37 is an outlier.

Another important application of quartiles is to represent data sets using box-and-whisker plots. A **box-and-whisker plot** is an exploratory data analysis tool that highlights the important features of a data set. To graph a box-and-whisker plot, you must know the following values.

PICTURING the WORLD

Of the first 43 U.S. presidents, Theodore Roosevelt was the youngest at the time of inauguration, at the age of 42. Ronald Reagan was the oldest president, inaugurated at the age of 69. The box-and-whisker plot summarizes the ages of the first 43 U.S. presidents at inauguration. (Source: whitehouse.gov)

Ages of U.S. Presidents at Inauguration



About how many U.S. presidents' ages are represented by the right whisker?

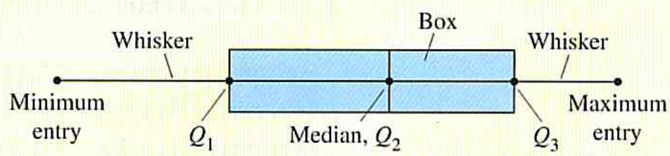
1. The minimum entry
2. The first quartile Q_1
3. The median Q_2
4. The third quartile Q_3
5. The maximum entry

These five numbers are called the **five-number summary** of the data set.

GUIDELINES

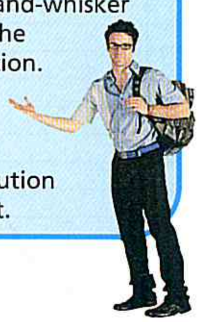
Drawing a Box-and-Whisker Plot

1. Find the five-number summary of the data set.
2. Construct a horizontal scale that spans the range of the data.
3. Plot the five numbers above the horizontal scale.
4. Draw a box above the horizontal scale from Q_1 to Q_3 and draw a vertical line in the box at Q_2.
5. Draw whiskers from the box to the minimum and maximum entries.



Insight

You can use a box-and-whisker plot to determine the shape of a distribution. Notice that the box-and-whisker plot in Example 4 represents a distribution that is skewed right.



EXAMPLE 4

Drawing a Box-and-Whisker Plot

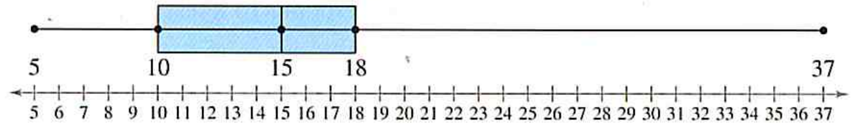
Draw a box-and-whisker plot that represents the 15 test scores given in Example 1. What can you conclude from the display?

See MINITAB and TI-83/84 steps on pages 124 and 125.

Solution The five-number summary of the test scores is below. Using these five numbers, you can construct the box-and-whisker plot shown.

Min = 5 $\quad Q_1 = 10 \quad Q_2 = 15 \quad Q_3 = 18 \quad$ Max = 37

Test Scores in CPR Class



Interpretation You can make several conclusions from the display. One is that about half the scores are between 10 and 18. By looking at the length of the right whisker, you can also conclude that the score of 37 is a possible outlier.

▸ Try It Yourself 4

Draw a box-and-whisker plot that represents the number of touchdowns scored by all Division 1A football teams listed in the chapter opener on page 39.

a. Find the *five-number summary* of the data set.
b. Construct a *horizontal scale* and *plot* the five numbers above it.
c. Draw the *box*, the *vertical line*, and the *whiskers*.
d. Make some conclusions.

Answer: Page A36

▸ Percentiles and Other Fractiles

Insight

Notice that the 25th percentile is the same as Q_1; the 50th percentile is the same as Q_2, or the median; the 75th percentile is the same as Q_3.

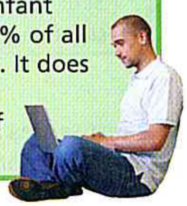
In addition to using quartiles to specify a measure of position, you can also use percentiles and deciles. These common fractiles are summarized as follows.

Fractiles	Summary	Symbols
Quartiles	Divide a data set into 4 equal parts.	Q_1, Q_2, Q_3
Deciles	Divide a data set into 10 equal parts.	$D_1, D_2, D_3, \ldots, D_9$
Percentiles	Divide a data set into 100 equal parts.	$P_1, P_2, P_3, \ldots, P_{99}$

Percentiles are often used in education and health-related fields to indicate how one individual compares with others in a group. They can also be used to identify unusually high or unusually low values. For instance, test scores and children's growth measurements are often expressed in percentiles. Scores or measurements in the 95th percentile and above are unusually high, while those in the 5th percentile and below are unusually low.

Study Tip

It is important that you understand what a percentile means. For instance, if the weight of a six-month-old infant is at the 78th percentile, the infant weighs more than 78% of all six-month-old infants. It does not mean that the infant weighs 78% of some ideal weight.

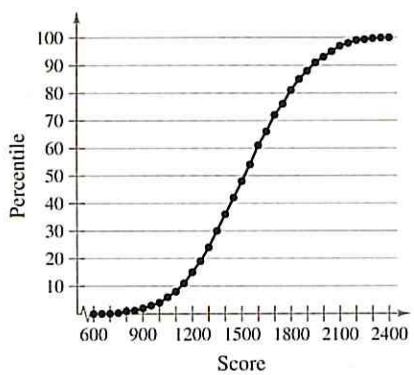


EXAMPLE 5

Interpreting Percentiles

The ogive represents the cumulative frequency distribution for SAT test scores of college-bound students in a recent year. What test score represents the 72nd percentile? How should you interpret this? *(Source: College Board Online)*

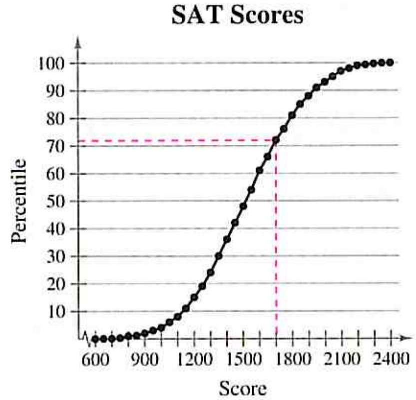
SAT Scores



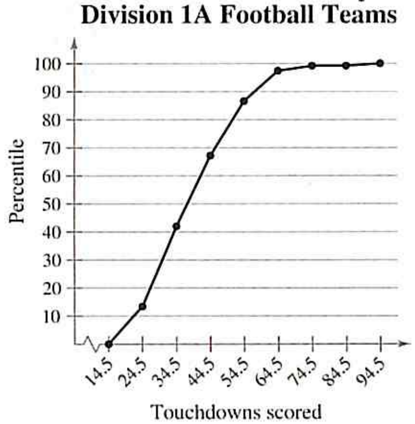
Solution From the ogive, you can see that the 72nd percentile corresponds to a test score of 1700.

Interpretation This means that 72% of the students had an SAT score of 1700 or less.

SAT Scores



Touchdowns Scored by Division 1A Football Teams



▸ Try It Yourself 5

The number of touchdowns scored by all Division 1A football teams are represented in the cumulative frequency graph at the left. At what percentile is a team that scores 40 touchdowns?

a. *Use the graph* to find the percentile that corresponds to the given touchdowns scored.

b. *Interpret* the results in the context of the data.

Answer: Page A36

The Standard Score

When you know the mean and standard deviation of a data set, you can measure a data value's position in the data set with a standard score, or z-score.

DEFINITION

The **standard score,** or ***z*-score,** represents the number of standard deviations a given value x falls from the mean μ. To find the z-score for a given value, use the following formula.

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

A z-score can be negative, positive, or zero. If z is negative, the corresponding x-value is below the mean. If z is positive, the corresponding x-value is above the mean. And if $z = 0$, the corresponding x-value is equal to the mean. A z-score can be used to identify an unusual value of a data set that is approximately bell-shaped.

EXAMPLE 6

Finding *z*-Scores

The mean speed of vehicles along a stretch of highway is 56 miles per hour with a standard deviation of 4 miles per hour. You measure the speed of three cars traveling along this stretch of highway as 62 miles per hour, 47 miles per hour, and 56 miles per hour. Find the z-score that corresponds to each speed. What can you conclude?

Solution The z-score that corresponds to each speed is calculated below.

$x = 62$ mph $\quad z = \frac{62 - 56}{4} = 1.5$

$x = 47$ mph $\quad z = \frac{47 - 56}{4} = -2.25$

$x = 56$ mph $\quad z = \frac{56 - 56}{4} = 0$

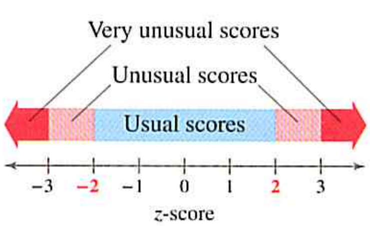
Interpretation From the z-scores, you can conclude that a speed of 62 miles per hour is 1.5 standard deviations above the mean; a speed of 47 miles per hour is 2.25 standard deviations below the mean; and a speed of 56 miles per hour is equal to the mean. If the distribution of the speeds is approximately bell-shaped, the car traveling 47 miles per hour is said to be traveling unusually slowly, because its speed corresponds to a z-score of -2.25.

▸ Try It Yourself 6

The monthly utility bills in a city have a mean of \$70 and a standard deviation of \$8. Find the z-scores that correspond to utility bills of \$60, \$71, and \$92. What can you conclude?

a. *Identify* μ and σ of the nonstandard normal distribution.
b. *Transform* each value to a z-score.
c. *Interpret* the results.

Answer: Page A36



When a distribution is approximately bell shaped, you know from the Empirical Rule that about 95% of the data lie within 2 standard deviations of the mean. So, when this distribution's values are transformed to z-scores, about 95% of the z-scores should fall between -2 and 2. A z-score outside of this range will occur about 5% of the time and would be considered unusual. So, according to the Empirical Rule, a z-score less than -3 or greater than 3 would be *very* unusual, with such a score occurring about 0.3% of the time.

In Example 6, you used z-scores to compare data values within the same data set. You can also use z-scores to compare data values from different data sets.

EXAMPLE 7

Comparing z-Scores from Different Data Sets

In 2007, Forest Whitaker won the Best Actor Oscar at age 45 for his role in the movie *The Last King of Scotland*. Helen Mirren won the Best Actress Oscar at age 61 for her role in *The Queen*. The mean age of all best actor winners is 43.7, with a standard deviation of 8.8. The mean age of all best actress winners is 36, with a standard deviation of 11.5. Find the z-score that corresponds to the age for each actor or actress. Then compare your results.

Solution

The z-score that corresponds to the age of each actor or actress is calculated below.

Forest Whitaker

$$z = \frac{x - \mu}{\sigma} = \frac{45 - 43.7}{8.8} \approx 0.15$$

Helen Mirren

$$z = \frac{x - \mu}{\sigma} = \frac{61 - 36}{11.5} \approx 2.17$$

The age of Forest Whitaker is 0.15 standard deviation above the mean, and the age of Helen Mirren is 2.17 standard deviations above the mean.

Interpretation The z-score corresponding to the age of Helen Mirren is more than two standard deviations from the mean, so it is considered unusual. Compared to other Best Actress winners, she is relatively older, whereas the age of Forest Whitaker is only slightly higher than the average age of other Best Actor winners.

▸ Try It Yourself 7

In 2007, Alan Arkin won the Best Supporting Actor Oscar at age 72 for his role in the movie *Little Miss Sunshine*. Jennifer Hudson won the Best Supporting Actress Oscar at age 25 for her role in *Dreamgirls*. The mean age of all best supporting actor winners is 50.1, with a standard deviation of 13.9. The mean age of all best supporting actress winners is 39.7, with a standard deviation of 14. Find the z-score that corresponds to the age for each actor or actress. Then compare your results.

a. *Identify* μ and σ of each nonstandard normal distribution.
b. *Transform* each value to a z-score.
c. *Compare* your results.

Answer: Page A36

2.5 EXERCISES

Building Basic Skills and Vocabulary

In Exercises 1 and 2, (a) find the three quartiles and (b) draw a box-and-whisker plot of the data.

1. 4 7 7 5 2 9 7 6 8 5 8 4 1 5 2 8 7 6 6 9

2. 2 7 1 3 1 2 8 9 9 2 5 4
7 3 7 5 4 7 2 3 5 9 5 6
3 9 3 4 9 8 8 2 3 9 5

3. The goals scored per game by a soccer team represent the first quartile for all teams in a league. What can you conclude about the team's goals scored per game?

4. A salesperson at a company sold $6,903,435 of hardware equipment last year, a figure that represented the eighth decile of sales performance at the company. What can you conclude about the salesperson's performance?

5. A student's score on an actuarial exam is in the 78th percentile. What can you conclude about the student's exam score?

6. A counselor tells a child's parents that their child's IQ is in the 93rd percentile for the child's age group. What can you conclude about the child's IQ?

True or False? *In Exercises 7–10, determine whether the statement is true or false. If it is false, rewrite it as a true statement.*

7. The second quartile is the median of an ordered data set.

8. The five numbers you need to graph a box-and-whisker plot are the minimum, the maximum, Q_1, Q_3, and the mean.

9. The 50th percentile is equivalent to Q_1.

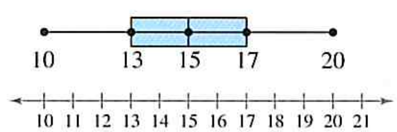
10. It is impossible to have a z-score of 0.

Using and Interpreting Concepts

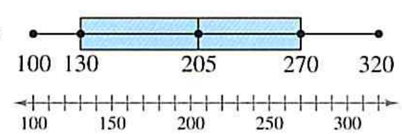
Graphical Analysis *In Exercises 11–16, use the box-and-whisker plot to identify*

(a) the minimum entry.
(b) the maximum entry.
(c) the first quartile.
(d) the second quartile.
(e) the third quartile.
(f) the interquartile range.

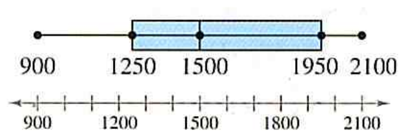
11.



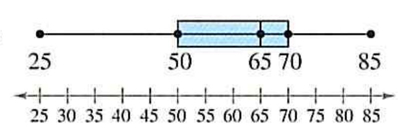
12.



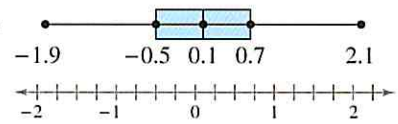
13.



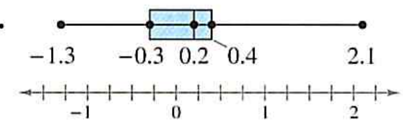
14.



15.

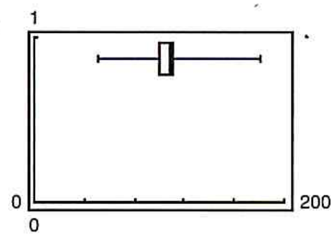


16.

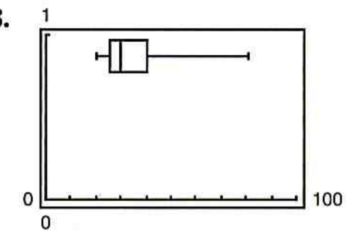


Interpreting Graphs *In Exercises 17–20, use the box-and-whisker plot to determine if the shape of the distribution represented is symmetric, skewed left, skewed right, or none of these. Justify your answer.*

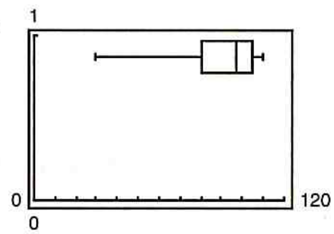
17.



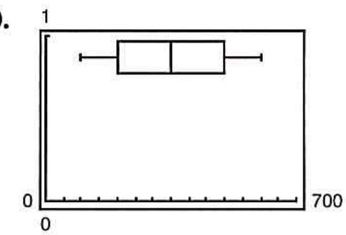
18.



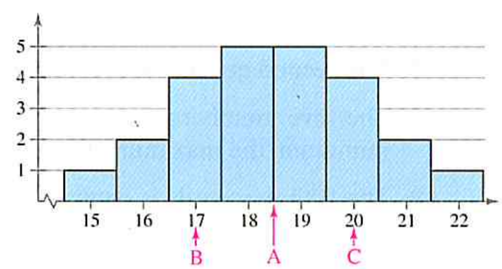
19.



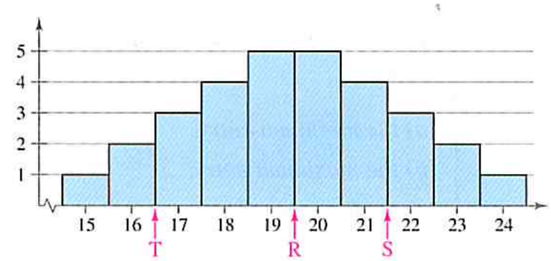
20.



21. Graphical Analysis The letters A, B, and C are marked on the histogram. Match them to Q_1, Q_2 (the median), and Q_3. Justify your answer.



22. Graphical Analysis The letters R, S, and T are marked on the histogram. Match them to P_{10}, P_{50}, and P_{80}. Justify your answer.



Using Technology to Find Quartiles and Draw Graphs *In Exercises 23–26, use a calculator or a computer to (a) find the data set's first, second, and third quartiles, and (b) draw a box-and-whisker plot that represents the data set.*

23. TV Viewing The number of hours of television watched per day by a sample of 28 people

2	4	1	5	7	2	5	4	4	2	3	6	4	3
5	2	0	3	5	9	4	5	2	1	3	6	7	2

24. Vacation Days The number of vacation days used by a sample of 20 employees in a recent year

3	9	2	1	7	5	3	2	2	6
4	0	10	0	3	5	7	8	6	5

25. Airplane Distances The distances (in miles) from an airport of a sample of 22 inbound and outbound airplanes

2.8	2.0	3.0	3.0	3.2	5.9	3.5	3.6
1.8	5.5	3.7	5.2	3.8	3.9	6.0	2.5
4.0	4.1	4.6	5.0	5.5	6.0		

26. Hourly Earnings The hourly earnings (in dollars) of a sample of 25 railroad equipment manufacturers

15.60	18.75	14.60	15.80	14.35	13.90	17.50	17.55	13.80
14.20	19.05	15.35	15.20	19.45	15.95	16.50	16,30	15.25
15.05	19.10	15.20	16.22	17.75	18.40	15.25		

27. TV Viewing Refer to the data set given in Exercise 23 and the box-and-whisker plot you drew that represents the data set.

(a) About 75% of the people watched no more than how many hours of television per day?

(b) What percent of the people watched more than 4 hours of television per day?

(c) If you randomly selected one person from the sample, what is the likelihood that the person watched less than 2 hours of television per day? Write your answer as a percent.

28. Manufacturer Earnings Refer to the data set given in Exercise 26 and the box-and-whisker plot you drew that represents the data set.

(a) About 75% of the manufacturers made less than what amount per hour?

(b) What percent of the manufacturers made more than $15.80 per hour?

(c) If you randomly selected one manufacturer from the sample, what is the likelihood that the manufacturer made less than $15.80 per hour? Write your answer as a percent.

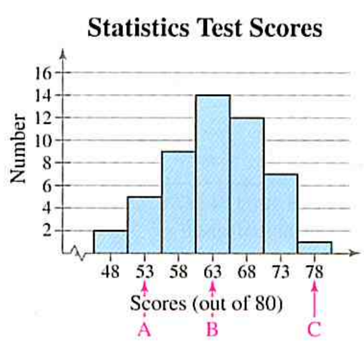
Graphical Analysis

In Exercises 29 and 30, the midpoints A, B, and C are marked on the histogram. Match them to the indicated z-scores. Which z-scores, if any, would be considered unusual?

29. $z = 0$

$z = 2.14$

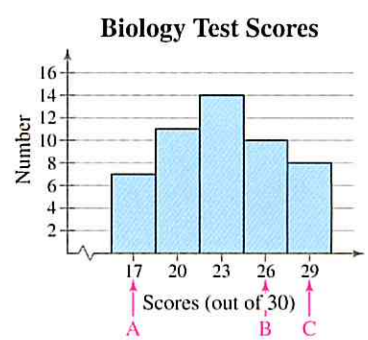
$z = -1.43$



30. $z = 0.77$

$z = 1.54$

$z = -1.54$



Comparing Test Scores *For the statistics test scores in Exercise 29, the mean is 63 and the standard deviation is 7.0, and for the biology test scores in Exercise 30 the mean is 23 and the standard deviation is 3.9. In Exercises 31–34, you are given the test scores of a student who took both tests.*

(a) Transform each test score to a z-score.

(b) Determine on which test the student had a better score.

31. A student gets a 73 on the statistics test and a 26 on the biology test.

32. A student gets a 60 on the statistics test and a 20 on the biology test.

33. A student gets a 78 on the statistics test and a 29 on the biology test.

34. A student gets a 63 on the statistics test and a 23 on the biology test.

35. Life Spans of Tires A certain brand of automobile tire has a mean life span of 35,000 miles and a standard deviation of 2250 miles. (Assume the life spans of the tires have a bell-shaped distribution.)

(a) The life spans of three randomly selected tires are 34,000 miles, 37,000 miles, and 31,000 miles. Find the z-score that corresponds to each life span. According to the z-scores, would the life spans of any of these tires be considered unusual?

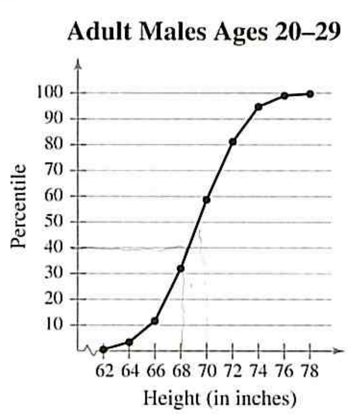
(b) The life spans of three randomly selected tires are 30,500 miles, 37,250 miles, and 35,000 miles. Using the Empirical Rule, find the percentile that corresponds to each life span.

36. Life Spans of Fruit Flies The life spans of a species of fruit fly have a bell-shaped distribution, with a mean of 33 days and a standard deviation of 4 days.

(a) The life spans of three randomly selected fruit flies are 34 days, 30 days, and 42 days. Find the z-score that corresponds to each life span and determine if any of these life spans are unusual.

(b) The life spans of three randomly selected fruit flies are 29 days, 41 days, and 25 days. Using the Empirical Rule, find the percentile that corresponds to each life span.

Interpreting Percentiles *In Exercises 37–42, use the cumulative frequency distribution to answer the questions. The cumulative frequency distribution represents the heights of males in the United States in the 20–29 age group. The heights have a bell-shaped distribution (see Picturing the World, page 88) with a mean of 69.6 inches and a standard deviation of 3.0 inches.* (Source: National Center for Health Statistics)



37. What height represents the 40th percentile? How should you interpret this?

38. What percentile is a height of 76 inches? How should you interpret this?

39. Three adult males in the 20–29 age group are randomly selected. Their heights are 74 inches, 62 inches, and 80 inches. Use z-scores to determine which heights, if any, are unusual.

40. Three adult males in the 20–29 age group are randomly selected. Their heights are 70 inches, 66 inches, and 68 inches. Use z-scores to determine which heights, if any, are unusual.

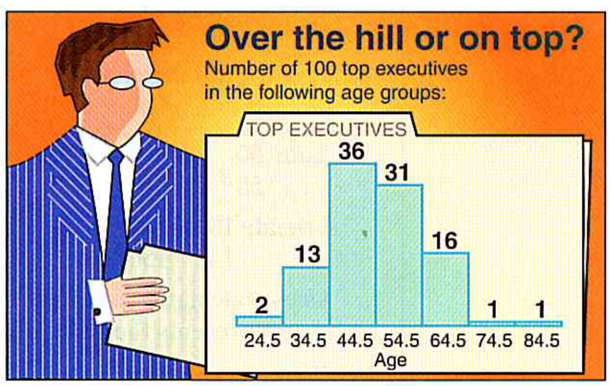
41. Find the z-score for a male in the 20–29 age group whose height is 71.1 inches. What percentile is this?

42. Find the z-score for a male in the 20–29 age group whose height is 66.3 inches. What percentile is this?

Extending Concepts

43. Ages of Executives The ages of a sample of 100 executives are listed.

31	62	51	44	61	47	49	45	40	52
60	51	67	47	63	54	59	43	63	52
50	54	61	41	48	49	51	54	39	54
47	52	36	53	74	33	53	68	44	40
60	42	50	48	42	42	36	57	42	48
56	51	54	42	27	43	43	41	54	49
49	47	51	28	54	36	36	41	60	55
42	59	35	65	48	56	82	39	54	49
61	56	57	32	38	48	64	51	45	46
62	63	59	63	32	47	40	37	49	57



(a) Order the data and find the first, second, and third quartiles.

(b) Draw a box-and-whisker plot that represents the data set.

(c) Interpret the results in the context of the data.

(d) On the basis of this sample, at what age would you expect to be an executive? Explain your reasoning.

(e) Which age groups, if any, can be considered unusual? Explain your reasoning.

Midquartile *Another measure of position is called the* **midquartile**. *You can find the midquartile of a data set by using the following formula.*

$$\textit{Midquartile} = \frac{Q_1 + Q_3}{2}$$

In Exercises 44–47, find the midquartile of the given data set.

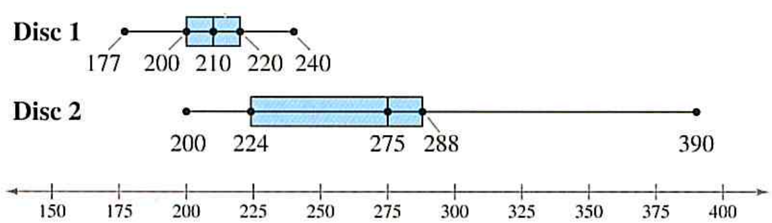
44. 5 7 1 2 3 10 8 7 5 3

45. 23 36 47 33 34 40 39 24 32 22 38 41

46. 12.3 9.7 8.0 15.4 16.1 11.8 12.7 13.4
12.2 8.1 7.9 10.3 11.2

47. 21.4 20.8 19.7 15.2 31.9 18.7 15.6 16.7
19.8 13.4 22.9 28.7 19.8 17.2 30.1

48. Song Lengths **Side-by-side box-and-whisker plots** can be used to compare two or more different data sets. Each box-and-whisker plot is drawn on the same number line to compare the data sets more easily. The song lengths (in seconds) from two different compact discs are given.



(a) Describe the shape of each distribution. Which disc has less variation in song lengths?

(b) Which distribution is more likely to have outliers? Explain your reasoning.

(c) Which disc do you think has a standard deviation of 16.3? Explain your reasoning.

49. Credit Card Purchases The monthly credit card purchases (rounded to the nearest dollar) over the last two years for you and a friend are listed.

You: 60 95 102 110 130 130 162 200 215 120 124 28
58 40 102 105 141 160 130 210 145 90 46 76

Friend: 100 125 132 90 85 75 140 160 180 190 160 105
145 150 151 82 78 115 170 158 140 130 165 125

Use a calculator or a computer to draw a side-by-side box-and-whisker plot that represents the data sets. Then describe the shapes of the distributions.

Finding Percentiles *You can find the percentile that corresponds to a specific data value x by using the following formula, then rounding the result to the nearest whole number.*

$$\text{Percentile of } x = \frac{\text{number of data values less than } x}{\text{total number of data values}} \cdot 100$$

In Exercises 50 and 51, use the information from Example 7 and the fact that there have been 80 Best Actor Oscars and 80 Best Actress Oscars awarded.

50. Fifty-one winners were younger than Forest Whitaker when they won the Best Actor Oscar. Find the percentile that corresponds to Forest Whitaker's age.

51. Only four winners were older than Helen Mirren when they won the Best Actress Oscar. Find the percentile that corresponds to Helen Mirren's age.

Uses & Abuses

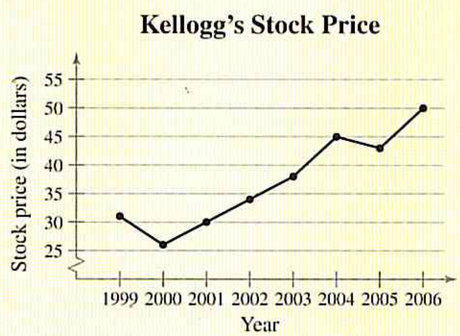
Uses

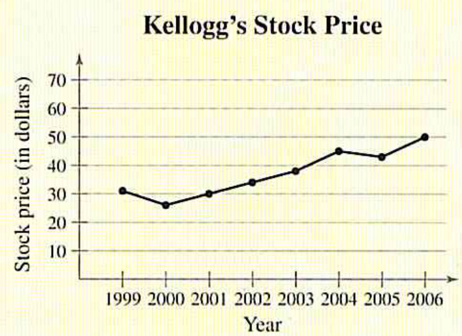
Descriptive statistics helps you see trends or patterns from a set of raw data. A good description of a data set consists of (1) a measure of the center of the data, (2) a measure of the variability (or spread) of the data, and (3) the shape (or distribution) of the data. When you read reports, news items, or advertisements prepared by other people, you are seldom given the raw data used for a study. Instead you see graphs, measures of central tendency, and measures of variability. To be a discerning reader, you need to understand the terms and techniques of descriptive statistics.

Abuses

Knowing how statistics are calculated can help you analyze questionable statistics. For example, suppose you are interviewing for a sales position and the company reports that the average yearly commission earned by the five people in its sales force is $60,000. This is a misleading statement if it is based on four commissions of $25,000 and one of $200,000. The median would more accurately describe the yearly commission, but they used the mean because it is a greater amount.

Statistical graphs can also be misleading. Compare the two time series charts below that show the year-end stock prices for Kellogg Company. The data are the same for each. The first graph, however, has a cropped vertical axis, which makes it appear that the stock price has increased greatly from 1999 to 2006. In the second graph, the scale on the vertical axis begins at zero. This graph correctly shows that the stock prices increased only modestly during this time period. *(Source: Kellogg Company)*





Ethics

Mark Twain helped popularize the saying, "There are three kinds of lies: lies, damned lies, and statistics." In short, even the most accurate statistics can be used to support studies or statements that are incorrect. Unscrupulous people can use misleading statistics to "prove" their point. Being informed about how statistics are calculated and questioning the data are ways to avoid being misled.

EXERCISES

1. In a newspaper or magazine, find an example of a graph that might lead to incorrect conclusions.
2. Describe a situation in which a statistic could be used to make a conclusion misleading.

2 CHAPTER SUMMARY

What did you learn?	EXAMPLE(S)	REVIEW EXERCISES
Section 2.1		
■ How to construct a frequency distribution including limits, midpoints, relative frequencies, cumulative frequencies, and boundaries	*1, 2*	*1*
■ How to construct frequency histograms, frequency polygons, relative frequency histograms, and ogives	*3–7*	*2–6*
Section 2.2		
■ How to graph quantitative data sets using the exploratory data analysis tools of stem-and-leaf plots and dot plots	*1–3*	*7, 8*
■ How to graph and interpret paired data sets using scatter plots and time series charts	*6, 7*	*9, 10*
■ How to graph qualitative data sets using pie charts and Pareto charts	*4, 5*	*11, 12*
Section 2.3		
■ How to find the mean, median, and mode of a population and a sample $\mu = \frac{\Sigma x}{N}, \bar{x} = \frac{\Sigma x}{n}$	*1–6*	*13, 14*
■ How to find a weighted mean of a data set and the mean of a frequency distribution $\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w}, \bar{x} = \frac{\Sigma(x \cdot f)}{n}$	*7, 8*	*15–18*
■ How to describe the shape of a distribution as symmetric, uniform, or skewed and how to compare the mean and median for each		*19–24*
Section 2.4		
■ How to find the range of a data set	*1*	*25, 26*
■ How to find the variance and standard deviation of a population and a sample $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}, s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$	*2–5*	*27–30*
■ How to use the Empirical Rule and Chebychev's Theorem to interpret standard deviation	*6–8*	*31–34*
■ How to approximate the sample standard deviation for grouped data $s = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{n-1}}$	*9, 10*	*35, 36*
Section 2.5		
■ How to find the quartiles and interquartile range of a data set	*1–3*	*37–39, 41*
■ How to draw a box-and-whisker plot	*4*	*40, 42*
■ How to interpret other fractiles such as percentiles	*5*	*43, 44*
■ How to find and interpret the standard score (z-score) $z = (x - \mu)/\sigma$	*6, 7*	*45–48*

2 REVIEW EXERCISES

Section 2.1

In Exercises 1 and 2, use the following data set. The data set represents the incomes (in thousands of dollars) of 20 employees at a small business.

30 28 26 39 34 33 20 39 28 33
26 39 32 28 31 39 33 31 33 32

1. Make a frequency distribution of the data set using five classes. Include the class midpoints, limits, boundaries, frequencies, relative frequencies, and cumulative frequencies.
2. Make a relative frequency histogram using the frequency distribution in Exercise 1. Then determine which class has the greatest relative frequency and which has the least relative frequency.

In Exercises 3 and 4, use the following data set. The data represent the actual liquid volumes (in ounces) in 24 twelve-ounce cans.

11.95 11.91 11.86 11.94 12.00 11.93 12.00 11.94
12.10 11.95 11.99 11.94 11.89 12.01 11.99 11.94
11.92 11.98 11.88 11.94 11.98 11.92 11.95 11.93

3. Make a frequency histogram using seven classes.
4. Make a relative frequency histogram of the data set using seven classes.

In Exercises 5 and 6, use the following data set. The data represent the number of rooms reserved during one night's business at a sample of hotels.

153 104 118 166 89 104 100 79
93 96 116 94 140 84 81 96
108 111 87 126 101 111 122 108
126 93 108 87 103 95 129 93

5. Make a frequency distribution with six classes and draw a frequency polygon.
6. Make an ogive of the data set using six classes.

Section 2.2

In Exercises 7 and 8, use the following data set. The data represent the average daily high temperatures (in degrees Fahrenheit) during the month of January for Chicago, Illinois. (Source: National Oceanic and Atmospheric Administration)

33 31 25 22 38 51 32 23
23 34 44 43 47 37 29 25
28 35 21 24 20 19 23 27
24 13 18 28 17 25 31

7. Make a stem-and-leaf plot of the data set. Use one line per stem.
8. Make a dot plot of the data set.
9. The following are the heights (in feet) and the number of stories of nine notable buildings in Miami. Use the data to construct a scatter plot. What type of pattern is shown in the scatter plot? *(Source: Emporis Buildings)*

Height (in feet)	764	625	520	510	484	492	450	430	410
Number of stories	55	47	51	28	34	39	33	31	40

10. The U.S. unemployment rate over a 12-year period is given. Use the data to construct a time series chart. *(Source: U.S. Bureau of Labor Statistics)*

Year	1995	1996	1997	1998	1999	2000
Unemployment rate	5.6	5.4	4.9	4.5	4.2	4.0

Year	2001	2002	2003	2004	2005	2006
Unemployment rate	4.7	5.8	6.0	5.5	5.1	4.6

In Exercises 11 and 12, use the following data set. The data set represents the top seven American Kennel Club registrations (in thousands) in 2006. (Source: American Kennel Club)

Breed	Labrador Retriever	Yorkshire Terrier	German Shepherd	Golden Retriever	Beagle	Dachshund	Boxer
Number registered (in thousands)	124	48	44	43	39	36	35

11. Make a Pareto chart of the data set.

12. Make a pie chart of the data set.

Section 2.3

13. Find the mean, median, and mode of the data set.

3 5 12 16 7 9 13 7 8 11

14. Find the mean, median, and mode of the data set.

42 36 39 42 44 45 42 42 36 38

15. Estimate the mean of the frequency distribution you made in Exercise 1.

16. The following frequency distribution shows the number of magazine subscriptions per household for a sample of 60 households. Find the mean number of subscriptions per household.

Number of magazines	0	1	2	3	4	5	6
Frequency	13	9	19	8	5	2	4

17. Six test scores are given. The first 5 test scores are 15% of the final grade, and the last test score is 25% of the final grade. Find the weighted mean of the test scores.

78 72 86 91 87 80

18. Four test scores are given. The first 3 test scores are 20% of the final grade, and the last test score is 40% of the final grade. Find the weighted mean of the test scores.

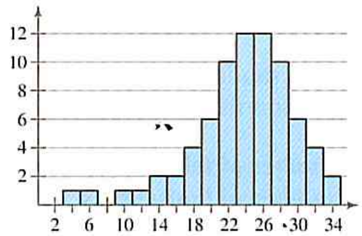
96 85 91 86

19. Describe the shape of the distribution in the histogram you made in Exercise 3. Is the distribution symmetric, uniform, or skewed?

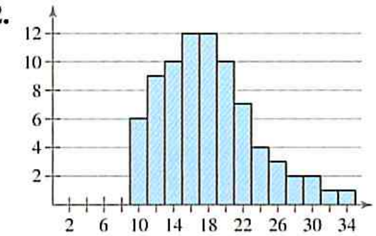
20. Describe the shape of the distribution in the histogram you made in Exercise 4. Is the distribution symmetric, uniform, or skewed?

In Exercises 21 and 22, determine whether the approximate shape of the distribution in the histogram is skewed right, skewed left, or symmetric.

21.



22.



23. For the histogram in Exercise 21, which is greater, the mean or the median?

24. For the histogram in Exercise 22, which is greater, the mean or the median?

Section 2.4

25. The data set represents the mean price of a movie ticket (in U.S. dollars) for a sample of 12 U.S. cities. Find the range of the data set.

7.82 7.38 6.42 6.76 6.34 7.44 6.15 5.46 7.92 6.58 8.26 7.17

26. The data set represents the mean price of a movie ticket (in U.S. dollars) for a sample of 12 Japanese cities. Find the range of the data set.

19.73 16.48 19.10 18.56 17.68 17.19
16.63 15.99 16.66 19.59 15.89 16.49

27. The mileage (in thousands) for a rental car company's fleet is listed. Find the population mean and standard deviation of the data.

4 2 9 12 15 3 6 8 1 4 14 12 3 3

28. The age of each Supreme Court justice as of March 19, 2007 is listed. Find the population mean and standard deviation of the data. *(Source: Supreme Court of the United States)*

52 86 71 70 67 58 74 68 56

29. Dormitory room prices (in dollars for one school year) for a sample of four-year universities are listed. Find the sample mean and the sample standard deviation of the data.

2445 2940 2399 1960 2421 2940 2657 2153
2430 2278 1947 2383 2710 2761 2377

30. Sample salaries (in dollars) of high school teachers are listed. Find the sample mean and standard deviation of the data.

49,632 54,619 58,298 48,250 51,842 50,875 53,219 49,924

31. The mean rate for satellite television from a sample of households was \$49.00 per month, with a standard deviation of \$2.50 per month. Between what two values do 99.7% of the data lie? (Assume a bell-shaped distribution.)

32. The mean rate for satellite television from a sample of households was \$49.50 per month, with a standard deviation of \$2.75 per month. Estimate the percent of satellite television rates between \$46.75 and \$52.25. (Assume that the data set has a bell-shaped distribution.)

33. The mean sale per customer for 40 customers at a gas station is \$36.00, with a standard deviation of \$8.00. On the basis of Chebychev's Theorem, at least how many of the customers spent between \$20.00 and \$52.00?

34. The mean length of the first 20 space shuttle flights was about 7 days, and the standard deviation was about 2 days. On the basis of Chebychev's Theorem, at least how many of the flights lasted between 3 days and 11 days? *(Source: NASA)*

35. From a random sample of households, the number of television sets are listed. Find the sample mean and standard deviation of the data.

Number of televisions	0	1	2	3	4	5
Number of households	1	8	13	10	5	3

36. From a random sample of airplanes, the number of defects found in their fuselages are listed. Find the sample mean and standard deviation of the data.

Number of defects	0	1	2	3	4	5	6
Number of airplanes	4	5	2	9	1	3	1

Section 2.5

In Exercises 37–40, use the following data set. The data represent the heights (in inches) of students in a statistics class.

52 54 55 56 56 56 58 59 60 61
61 63 65 67 68 68 70 71 72

37. Find the height that corresponds to the first quartile.

38. Find the height that corresponds to the third quartile.

39. Find the interquartile range.

40. Make a box-and-whisker plot of the data.

41. Find the interquartile range of the data from Exercise 14.

42. The weights (in pounds) of the defensive players on a high school football team are given. Make a box-and-whisker plot of the data.

173 145 205 192 197 227 156 240 172 185
208 185 190 167 212 228 190 184 195

43. A student's test grade of 68 represents the 77th percentile of the grades. What percent of students scored higher than 68?

44. In 2007 there were 768 "oldies" radio stations in the United States. If one station finds that 84 stations have a larger daily audience than it has, what percentile does this station come closest to in the daily audience rankings? *(Source: Radio-locator.com)*

In Exercises 45–48, use the following information. The weights of 19 high school football players have a bell-shaped distribution, with a mean of 186 pounds and a standard deviation of 18 pounds. Use z-scores to determine if the weights of the following randomly selected football players are unusual.

45. 213 pounds

46. 141 pounds

47. 178 pounds

48. 249 pounds

2 CHAPTER QUIZ

Take this quiz as you would take a quiz in class. After you are done, check your work against the answers given in the back of the book.

1. The data set is the number of minutes a sample of 25 people exercise each week.

108	139	120	123	120	132	123	131	131
157	150	124	111	101	135	119	116	117
127	128	139	119	118	114	127		

(a) Make a frequency distribution of the data set using five classes. Include class limits, midpoints, frequencies, boundaries, relative frequencies, and cumulative frequencies.

(b) Display the data using a frequency histogram and a frequency polygon on the same axes.

(c) Display the data using a relative frequency histogram.

(d) Describe the distribution's shape as symmetric, uniform, or skewed.

(e) Display the data using a box-and-whisker plot.

(f) Display the data using an ogive.

2. Use frequency distribution formulas to approximate the sample mean and standard deviation of the data set in Exercise 1.

3. U.S. sporting goods sales (in billions of dollars) can be classified in four areas: clothing (11.7), footwear (15.7), equipment (24.0), and recreational transport (38.5). Display the data using (a) a pie chart and (b) a Pareto chart. *(Source: National Sporting Goods Association)*

4. Weekly salaries (in dollars) for a sample of registered nurses are listed.

774 446 1019 795 908 667 444 960

(a) Find the mean, the median, and the mode of the salaries. Which best describes a typical salary?

(b) Find the range, variance, and standard deviation of the data set. Interpret the results in the context of the real-life setting.

5. The mean price of new homes from a sample of houses is \$155,000 with a standard deviation of \$15,000. The data set has a bell-shaped distribution. Between what two prices do 95% of the houses fall?

6. Refer to the sample statistics from Exercise 5 and use z-scores to determine which, if any, of the following house prices is unusual.

(a) \$200,000 (b) \$55,000 (c) \$175,000 (d) \$122,000

7. The number of wins for each Major League Baseball team in 2006 are listed. *(Source: Major League Baseball)*

97	87	86	70	61	96	95	90	78	62
93	89	80	78	97	85	79	78	71	83
82	80	75	67	66	88	88	76	76	76

(a) Find the quartiles of the data set.

(b) Find the interquartile range.

(c) Draw a box-and-whisker plot.

Putting It All Together

REAL Statistics — Real Decisions

You are a consumer journalist for a newspaper. You have received several letters and e-mails from readers who are concerned about the cost of their automobile insurance premiums. One of the readers wrote the following:

> *"I think, on the average, a driver in our city pays a higher automobile insurance premium than drivers in other cities like ours in this state."*

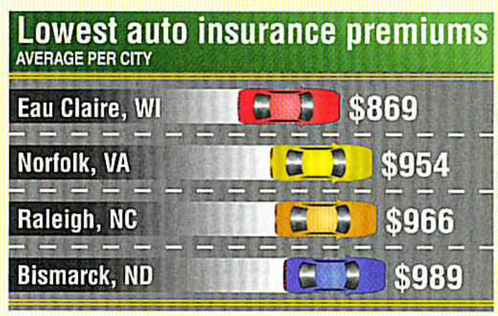
Your editor asks you to investigate the costs of insurance premiums and write an article about it. You have gathered the data shown at the right (your city is City A). The data represent the automobile insurance premiums paid annually (in dollars) by a random sample of drivers in your city and three other cities of similar size in your state. (The prices of the premiums from the sample include comprehensive, collision, bodily injury, property damage, and uninsured motorist coverage.)

runzheimer INTERNATIONAL

The Prices, in Dollars, of Automobile Insurance Premiums Paid by 10 Randomly Selected Drivers in 4 Cities

City A	City B	City C	City D
2465	2514	2030	2345
1984	1600	1450	2152
2545	1545	2715	1570
1640	2716	2145	1850
1983	1987	1600	1450
2302	2200	1430	1745
2542	2005	1545	1590
1875	1945	1792	1800
1920	1380	1645	2575
2655	2400	1368	2016

(Adapted from: Runzheimer International)



■ Exercises

1. *How Would You Do It?*

(a) How would you investigate the statement about the price of automobile insurance premiums?

(b) What statistical measures in this chapter would you use?

2. *Displaying the Data*

(a) What type of graph would you choose to display the data? Why?

(b) Construct the graph from part (a).

(c) On the basis of what you did in part (b), does it appear that the average automobile insurance premium in your city, City A, is higher than in any of the other cities? Explain.

3. *Measuring the Data*

(a) What statistical measures discussed in this chapter would you use to analyze the automobile insurance premium data?

(b) Calculate the measures from part (a).

(c) Compare the measures from part (b) with the graph you made in Exercise 2. Do the measurements support your conclusion in Exercise 2? Explain.

4. *Discussing the Data*

(a) What would you tell your readers? Is the average automobile insurance premium in your city more than in the other cities?

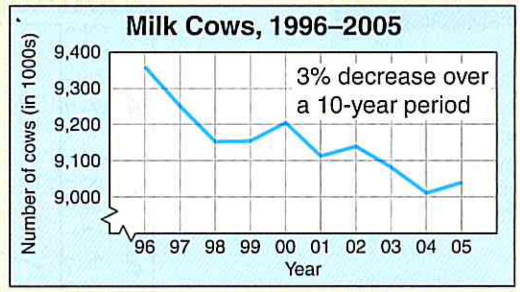
(b) What reasons might you give to your readers as to why the prices of automobile insurance premiums vary from city to city?

TECHNOLOGY

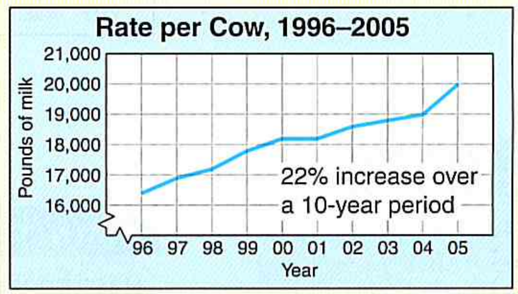
MINITAB | EXCEL | T1-83/84

Dairy Farmers of America is an association that provides help to dairy farmers. Part of this help is gathering and distributing statistics on milk production.

www.dfamilk.com



(Source: National Agricultural Statistics Service)



(Source: National Agricultural Statistics Service)

From 1996 to 2005, the number of dairy cows in the United States decreased and the yearly milk production increased.

MONTHLY MILK PRODUCTION

The following data set was supplied by a dairy farmer. It lists the monthly milk production (in pounds) for 50 Holstein dairy cows. *(Source: Matlink Dairy, Clymer, NY)*

2825	2072	2733	2069	2484
4285	2862	3353	1449	2029
1258	2982	2045	1677	1619
2597	3512	2444	1773	2284
1884	2359	2046	2364	2669
3109	2804	1658	2207	2159
2207	2882	1647	2051	2202
3223	2383	1732	2230	1147
2711	1874	1979	1319	2923
2281	1230	1665	1294	2936

EXERCISES

In Exercises 1–4, use a computer or calculator. If possible, print your results.

1. Find the sample mean of the data.
2. Find the sample standard deviation of the data.
3. Make a frequency distribution for the data. Use a class width of 500.
4. Draw a histogram for the data. Does the distribution appear to be bell shaped?
5. What percent of the distribution lies within one standard deviation of the mean? Within two standard deviations of the mean? How do these results agree with the Empirical Rule?

In Exercises 6–8, use the frequency distribution found in Exercise 3.

6. Use the frequency distribution to estimate the sample mean of the data. Compare your results with Exercise 1.
7. Use the frequency distribution to find the sample standard deviation for the data. Compare your results with Exercise 2.
8. **Writing** Use the results of Exercises 6 and 7 to write a general statement about the mean and standard deviation for grouped data. Do the formulas for grouped data give results that are as accurate as the individual entry formulas?

Extended solutions are given in the *Technology Supplement*.
Technical instruction is provided for MINITAB, Excel, and the TI-83/84.