2 CHAPTER

DESCRIPTIVE STATISTICS

Brothers Sam and Bud Walton opened the first Wal-Mart store in 1962. Today, the Walton family is one of the richest families in the world. Members of the Walton family held four spots in the top 50 richest people in the world in 2009.

WHERE YOU'VE BEEN

In Chapter 1, you learned that there are many ways to collect data. Usually, researchers must work with sample data in order to analyze populations, but occasionally it is possible to collect all the data for a given population. For instance, the following represents the ages of the 50 richest people in the world in 2009.

89, 89, 87, 86, 86, 85, 83, 83, 82, 81, 80, 78, 78, 77, 76, 73, 73, 73, 72, 69, 69, 68, 67, 66, 66, 65, 65, 64, 63, 61, 61, 60, 59, 58, 57, 56, 54, 54, 53, 53, 51, 51, 49, 47, 46, 44, 43, 42, 36, 35

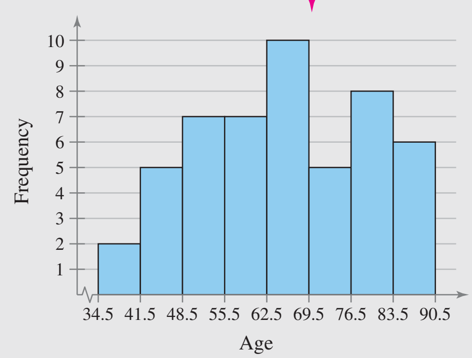
WHERE YOU'RE GOING

In Chapter 2, you will learn ways to organize and describe data sets. The goal is to make the data easier to understand by describing trends, averages, and variations. For instance, in the raw data showing the ages of the 50 richest people in the world in 2009, it is not easy to see any patterns or special characteristics. Here are some ways you can organize and describe the data.

Make a frequency distribution table.

Class	Frequency, f
35–41	2
42–48	5
49–55	7
56–62	7
63–69	10
70–76	5
77–83	8
84–90	6

Draw a histogram.



$$\text{Mean} = \frac{89 + 89 + 87 + 86 + 86 + \cdots + 43 + 42 + 36 + 35}{50}$$

$$= \frac{3263}{50}$$

$$= 65.26 \text{ years old}$$

Find an average.

$$\text{Range} = 89 - 35$$

$$= 54 \text{ years}$$

Find how the data vary.

2.1 Frequency Distributions and Their Graphs

WHAT YOU SHOULD LEARN

- How to construct a frequency distribution including limits, midpoints, relative frequencies, cumulative frequencies, and boundaries
- How to construct frequency histograms, frequency polygons, relative frequency histograms, and ogives

Frequency Distributions ▸ Graphs of Frequency Distributions

▸ FREQUENCY DISTRIBUTIONS

You will learn that there are many ways to organize and describe a data set. Important characteristics to look for when organizing and describing a data set are its **center,** its **variability** (or spread), and its **shape.** Measures of center and shapes of distributions are covered in Section 2.3.

When a data set has many entries, it can be difficult to see patterns. In this section, you will learn how to organize data sets by grouping the data into *intervals* called *classes* and forming a *frequency distribution*. You will also learn how to use frequency distributions to construct graphs.

DEFINITION

A **frequency distribution** is a table that shows **classes** or **intervals** of data entries with a count of the number of entries in each class. The **frequency** ***f*** of a class is the number of data entries in the class.

Example of a Frequency Distribution

Class	Frequency, f
1–5	5
6–10	8
11–15	6
16–20	8
21–25	5
26–30	4

In the frequency distribution shown at the left there are six classes. The frequencies for each of the six classes are 5, 8, 6, 8, 5, and 4. Each class has a **lower class limit,** which is the least number that can belong to the class, and an **upper class limit,** which is the greatest number that can belong to the class. In the frequency distribution shown, the lower class limits are 1, 6, 11, 16, 21, and 26, and the upper class limits are 5, 10, 15, 20, 25, and 30. The **class width** is the distance between lower (or upper) limits of consecutive classes. For instance, the class width in the frequency distribution shown is $6 - 1 = 5$.

The difference between the maximum and minimum data entries is called the **range.** In the frequency table shown, suppose the maximum data entry is 29, and the minimum data entry is 1. The range then is $29 - 1 = 28$. You will learn more about the range of a data set in Section 2.4.

STUDY TIP

In a frequency distribution, it is best if each class has the same width. Answers shown will use the minimum data value for the lower limit of the first class. Sometimes it may be more convenient to choose a lower limit that is slightly lower than the minimum value. The frequency distribution produced will vary slightly.

GUIDELINES

Constructing a Frequency Distribution from a Data Set

1. Decide on the number of classes to include in the frequency distribution. The number of classes should be between 5 and 20; otherwise, it may be difficult to detect any patterns.
2. Find the class width as follows. Determine the range of the data, divide the range by the number of classes, and *round up to the next convenient number*.
3. Find the class limits. You can use the minimum data entry as the lower limit of the first class. To find the remaining lower limits, add the class width to the lower limit of the preceding class. Then find the upper limit of the first class. Remember that classes cannot overlap. Find the remaining upper class limits.
4. Make a tally mark for each data entry in the row of the appropriate class.
5. Count the tally marks to find the total frequency f for each class.

EXAMPLE 1

▸ Constructing a Frequency Distribution from a Data Set

The following sample data set lists the prices (in dollars) of 30 portable global positioning system (GPS) navigators. Construct a frequency distribution that has seven classes.

90	130	400	200	350	70	325	250	150	250
275	270	150	130	59	200	160	450	300	130
220	100	200	400	200	250	95	180	170	150

INSIGHT

If you obtain a whole number when calculating the class width of a frequency distribution, use the next whole number as the class width. Doing this ensures that you will have enough space in your frequency distribution for all the data values.

▸ Solution

1. The number of classes (7) is stated in the problem.
2. The minimum data entry is 59 and the maximum data entry is 450, so the range is $450 - 59 = 391$. Divide the range by the number of classes and round up to find the class width.

$$\text{Class width} = \frac{391}{7} \qquad \frac{\text{Range}}{\text{Number of classes}}$$

$$\approx 55.86 \qquad \text{Round up to 56.}$$

3. The minimum data entry is a convenient lower limit for the first class. To find the lower limits of the remaining six classes, add the class width of 56 to the lower limit of each previous class. The upper limit of the first class is 114, which is one less than the lower limit of the second class. The upper limits of the other classes are $114 + 56 = 170$, $170 + 56 = 226$, and so on. The lower and upper limits for all seven classes are shown.

Lower limit	Upper limit
59	114
115	170
171	226
227	282
283	338
339	394
395	450

4. Make a tally mark for each data entry in the appropriate class. For instance, the data entry 130 is in the 115–170 class, so make a tally mark in that class. Continue until you have made a tally mark for each of the 30 data entries.
5. The number of tally marks for a class is the frequency of that class.

The frequency distribution is shown in the following table. The first class, 59–114, has five tally marks. So, the frequency of this class is 5. Notice that the sum of the frequencies is 30, which is the number of entries in the sample data set. The sum is denoted by Σf, where Σ is the uppercase Greek letter **sigma.**

STUDY TIP

The uppercase Greek letter sigma (Σ) is used throughout statistics to indicate a summation of values.

Frequency Distribution for Prices (in dollars) of GPS Navigators

Prices

Number of GPS navigators

Class	Tally	Frequency, f
59–114	卌	5
115–170	卌 |||	8
171–226	卌 |	6
227–282	卌	5
283–338	||	2
339–394	|	1
395–450	|||	3
		$\Sigma f = 30$

Check that the sum of the frequencies equals the number in the sample.

▸ Try It Yourself 1

Construct a frequency distribution using the ages of the 50 richest people data set listed in the Chapter Opener on page 37. Use eight classes.

a. State the *number* of classes.
b. Find the minimum and maximum values and the *class width*.
c. Find the *class limits*.
d. *Tally* the data entries.
e. Write the *frequency* f of each class.

Answer: Page A30

After constructing a standard frequency distribution such as the one in Example 1, you can include several additional features that will help provide a better understanding of the data. These features (the *midpoint*, *relative frequency*, and *cumulative frequency* of each class) can be included as additional columns in your table.

DEFINITION

The **midpoint** of a class is the sum of the lower and upper limits of the class divided by two. The midpoint is sometimes called the *class mark*.

$$\text{Midpoint} = \frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$$

The **relative frequency** of a class is the portion or percentage of the data that falls in that class. To find the relative frequency of a class, divide the frequency f by the sample size n.

$$\text{Relative frequency} = \frac{\text{Class frequency}}{\text{Sample size}} = \frac{f}{n}$$

The **cumulative frequency** of a class is the sum of the frequencies of that class and all previous classes. The cumulative frequency of the last class is equal to the sample size n.

After finding the first midpoint, you can find the remaining midpoints by adding the class width to the previous midpoint. For instance, if the first midpoint is 86.5 and the class width is 56, then the remaining midpoints are

$$86.5 + 56 = 142.5$$

$$142.5 + 56 = 198.5$$

$$198.5 + 56 = 254.5$$

$$254.5 + 56 = 310.5$$

and so on.

You can write the relative frequency as a fraction, decimal, or percent. The sum of the relative frequencies of all the classes should be equal to 1, or 100%. Due to rounding, the sum may be slightly less than or greater than 1. So, values such as 0.99 and 1.01 are sufficient.

EXAMPLE 2

Finding Midpoints, Relative Frequencies, and Cumulative Frequencies

Using the frequency distribution constructed in Example 1, find the midpoint, relative frequency, and cumulative frequency of each class. Identify any patterns.

Solution

The midpoints, relative frequencies, and cumulative frequencies of the first three classes are calculated as follows.

Class	f	**Midpoint**	**Relative frequency**	**Cumulative frequency**
59–114	5	$\frac{59 + 114}{2} = 86.5$	$\frac{5}{30} \approx 0.17$	5
115–170	8	$\frac{115 + 170}{2} = 142.5$	$\frac{8}{30} \approx 0.27$	$5 + 8 = 13$
171–226	6	$\frac{171 + 226}{2} = 198.5$	$\frac{6}{30} = 0.2$	$13 + 6 = 19$

The remaining midpoints, relative frequencies, and cumulative frequencies are shown in the following expanded frequency distribution.

Frequency Distribution for Prices (in dollars) of GPS Navigators

Prices → Class; Number of GPS navigators → Frequency, f; Portion of GPS navigators → Relative frequency

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency
59–114	5	86.5	0.17	5
115–170	8	142.5	0.27	13
171–226	6	198.5	0.2	19
227–282	5	254.5	0.17	24
283–338	2	310.5	0.07	26
339–394	1	366.5	0.03	27
395–450	3	422.5	0.1	30
	$\Sigma f = 30$		$\Sigma \frac{f}{n} \approx 1$	

Interpretation There are several patterns in the data set. For instance, the most common price range for GPS navigators was \$115 to \$170.

Try It Yourself 2

Using the frequency distribution constructed in Try It Yourself 1, find the midpoint, relative frequency, and cumulative frequency of each class. Identify any patterns.

a. Use the formulas to find each *midpoint*, *relative frequency*, and *cumulative frequency*.
b. *Organize* your results in a frequency distribution.
c. *Identify* patterns that emerge from the data.

Answer: Page A31

GRAPHS OF FREQUENCY DISTRIBUTIONS

Sometimes it is easier to identify patterns of a data set by looking at a graph of the frequency distribution. One such graph is a *frequency histogram*.

DEFINITION

A **frequency histogram** is a bar graph that represents the frequency distribution of a data set. A histogram has the following properties.

1. The horizontal scale is quantitative and measures the data values.
2. The vertical scale measures the frequencies of the classes.
3. Consecutive bars must touch.

Because consecutive bars of a histogram must touch, bars must begin and end at class boundaries instead of class limits. **Class boundaries** are the numbers that separate classes *without* forming gaps between them. If data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries. To find the upper class boundaries, add 0.5 to each upper limit. The upper boundary of a class will equal the lower boundary of the next higher class.

EXAMPLE 3 SC Report 2

Constructing a Frequency Histogram

Draw a frequency histogram for the frequency distribution in Example 2. Describe any patterns.

Solution

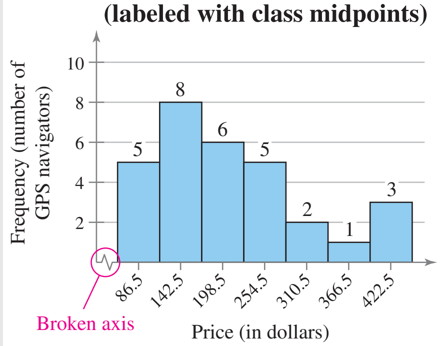
First, find the class boundaries. Because the data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries and add 0.5 to each upper limit to find the upper class boundaries. So, the lower and upper boundaries of the first class are as follows.

$$\text{First class lower boundary} = 59 - 0.5 = 58.5$$
$$\text{First class upper boundary} = 114 + 0.5 = 114.5$$

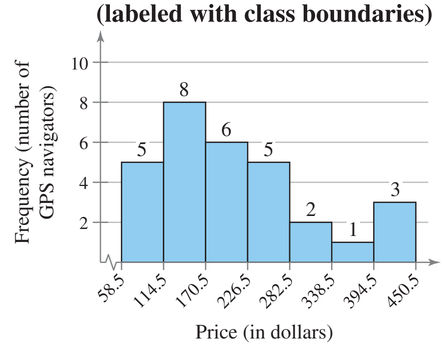
Class	Class boundaries	Frequency, f
59–114	58.5–114.5	5
115–170	114.5–170.5	8
171–226	170.5–226.5	6
227–282	226.5–282.5	5
283–338	282.5–338.5	2
339–394	338.5–394.5	1
395–450	394.5–450.5	3

The boundaries of the remaining classes are shown in the table. To construct the histogram, choose possible frequency values for the vertical scale. You can mark the horizontal scale either at the midpoints or at the class boundaries. Both histograms are shown.

Prices of GPS Navigators (labeled with class midpoints)



Prices of GPS Navigators (labeled with class boundaries)



Interpretation From either histogram, you can see that more than half of the GPS navigators are priced below $226.50.

INSIGHT

It is customary in bar graphs to have spaces between the bars, whereas with histograms, it is customary that the bars have no spaces between them.

Try It Yourself 3

Use the frequency distribution from Try It Yourself 2 to construct a frequency histogram that represents the ages of the 50 richest people. Describe any patterns.

a. Find the *class boundaries*.
b. Choose appropriate *horizontal and vertical scales*.
c. Use the frequency distribution to *find the height of each bar*.
d. *Describe* any patterns in the data.

Answer: Page A31

Another way to graph a frequency distribution is to use a frequency polygon. A **frequency polygon** is a line graph that emphasizes the continuous change in frequencies.

EXAMPLE 4

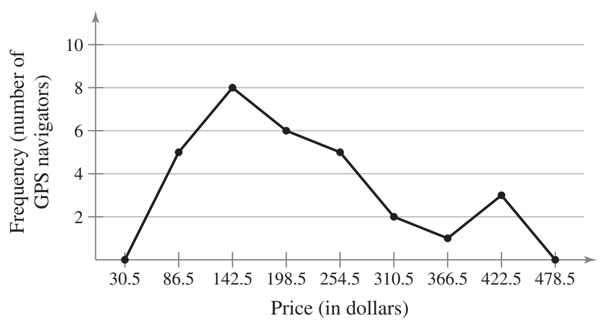
Constructing a Frequency Polygon

Draw a frequency polygon for the frequency distribution in Example 2. Describe any patterns.

Solution

To construct the frequency polygon, use the same horizontal and vertical scales that were used in the histogram labeled with class midpoints in Example 3. Then plot points that represent the midpoint and frequency of each class and connect the points in order from left to right. Because the graph should begin and end on the horizontal axis, extend the left side to one class width before the first class midpoint and extend the right side to one class width after the last class midpoint.

Prices of GPS Navigators



Interpretation You can see that the frequency of GPS navigators increases up to $142.50 and then decreases.

Try It Yourself 4

Use the frequency distribution from Try It Yourself 2 to construct a frequency polygon that represents the ages of the 50 richest people. Describe any patterns.

a. Choose appropriate *horizontal and vertical scales*.
b. *Plot points* that represent the midpoint and frequency of each class.
c. *Connect the points* and extend the sides as necessary.
d. *Describe* any patterns in the data.

Answer: Page A31

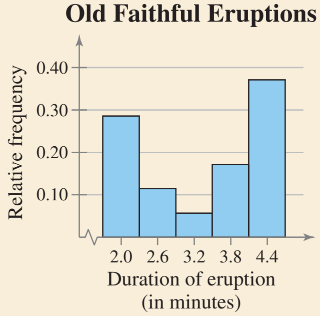
STUDY TIP

A histogram and its corresponding frequency polygon are often drawn together. If you have not already constructed the histogram, begin constructing the frequency polygon by choosing appropriate horizontal and vertical scales. The horizontal scale should consist of the class midpoints, and the vertical scale should consist of appropriate frequency values.

A **relative frequency histogram** has the same shape and the same horizontal scale as the corresponding frequency histogram. The difference is that the vertical scale measures the *relative* frequencies, not frequencies.

PICTURING THE WORLD

Old Faithful, a geyser at Yellowstone National Park, erupts on a regular basis. The time spans of a sample of eruptions are given in the relative frequency histogram. (Source: Yellowstone National Park)



Fifty percent of the eruptions last less than how many minutes?

EXAMPLE 5

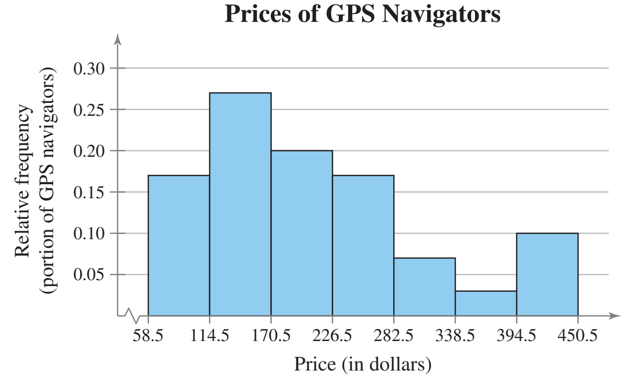
SC Report 3

Constructing a Relative Frequency Histogram

Draw a relative frequency histogram for the frequency distribution in Example 2.

Solution

The relative frequency histogram is shown. Notice that the shape of the histogram is the same as the shape of the frequency histogram constructed in Example 3. The only difference is that the vertical scale measures the relative frequencies.



Interpretation From this graph, you can quickly see that 0.27 or 27% of the GPS navigators are priced between \$114.50 and \$170.50, which is not as immediately obvious from the frequency histogram.

Try It Yourself 5

Use the frequency distribution in Try It Yourself 2 to construct a relative frequency histogram that represents the ages of the 50 richest people.

a. *Use the same horizontal scale* that was used in the frequency histogram in the Chapter Opener.
b. *Revise the vertical scale* to reflect relative frequencies.
c. Use the relative frequencies to *find the height of each bar*.

Answer: Page A31

If you want to describe the number of data entries that are equal to or below a certain value, you can easily do so by constructing a *cumulative frequency graph*.

DEFINITION

A **cumulative frequency graph,** or **ogive** (pronounced $\bar{o}'j\bar{\imath}ve$), is a line graph that displays the cumulative frequency of each class at its upper class boundary. The upper boundaries are marked on the horizontal axis, and the cumulative frequencies are marked on the vertical axis.

GUIDELINES

Constructing an Ogive (Cumulative Frequency Graph)

1. Construct a frequency distribution that includes cumulative frequencies as one of the columns.
2. Specify the horizontal and vertical scales. The horizontal scale consists of upper class boundaries, and the vertical scale measures cumulative frequencies.
3. Plot points that represent the upper class boundaries and their corresponding cumulative frequencies.
4. Connect the points in order from left to right.
5. The graph should start at the lower boundary of the first class (cumulative frequency is zero) and should end at the upper boundary of the last class (cumulative frequency is equal to the sample size).

EXAMPLE 6

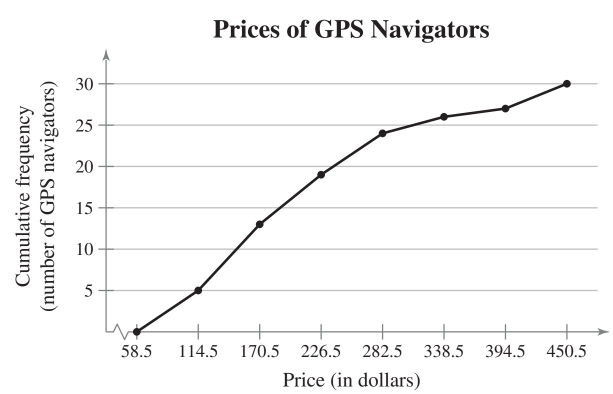
Constructing an Ogive

Draw an ogive for the frequency distribution in Example 2. Estimate how many GPS navigators cost $300 or less. Also, use the graph to estimate when the greatest increase in price occurs.

Solution

Using the cumulative frequencies, you can construct the ogive shown. The upper class boundaries, frequencies, and cumulative frequencies are shown in the table. Notice that the graph starts at 58.5, where the cumulative frequency is 0, and the graph ends at 450.5, where the cumulative frequency is 30.

Upper class boundary	*f*	Cumulative frequency
114.5	5	5
170.5	8	13
226.5	6	19
282.5	5	24
338.5	2	26
394.5	1	27
450.5	3	30



Interpretation From the ogive, you can see that about 25 GPS navigators cost $300 or less. It is evident that the greatest increase occurs between $114.50 and $170.50, because the line segment is steepest between these two class boundaries.

Another type of ogive uses percent as the vertical axis instead of frequency (see Example 5 in Section 2.5).

▸ Try It Yourself 6

Use the frequency distribution from Try It Yourself 2 to construct an ogive that represents the ages of the 50 richest people. Estimate the number of people who are 80 years old or younger.

a. Specify the *horizontal and vertical scales*.
b. *Plot* the points given by the upper class boundaries and the cumulative frequencies.
c. *Construct* the graph.
d. *Estimate* the number of people who are 80 years old or younger.
e. *Interpret* the results in the context of the data. *Answer: Page A31*

EXAMPLE 7

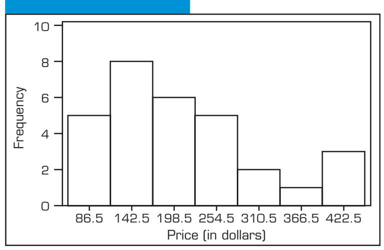
▸ Using Technology to Construct Histograms

Use a calculator or a computer to construct a histogram for the frequency distribution in Example 2.

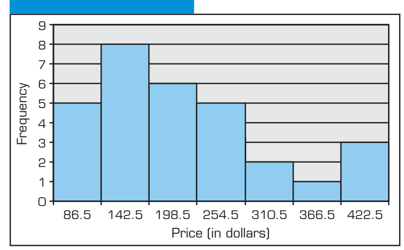
▸ Solution

MINITAB, Excel, and the TI-83/84 Plus each have features for graphing histograms. Try using this technology to draw the histograms as shown.

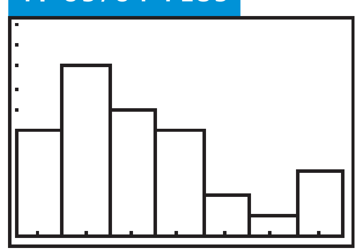
MINITAB



EXCEL



TI-83/84 PLUS



STUDY TIP

Detailed instructions for using MINITAB, Excel, and the TI-83/84 Plus are shown in the Technology Guide that accompanies this text. For instance, here are instructions for creating a histogram on a TI-83/84 Plus.

STAT ENTER

Enter midpoints in L1.
Enter frequencies in L2.

2nd STATPLOT

Turn on Plot 1.
Highlight Histogram.

Xlist: L1
Freq: L2

ZOOM 9

WINDOW

Xscl=56

GRAPH

▸ Try It Yourself 7

Use a calculator or a computer and the frequency distribution from Try It Yourself 2 to construct a frequency histogram that represents the ages of the 50 richest people.

a. *Enter* the data
b. *Construct* the histogram. *Answer: Page A31*

2.1 EXERCISES

BUILDING BASIC SKILLS AND VOCABULARY

1. What are some benefits of representing data sets using frequency distributions? What are some benefits of using graphs of frequency distributions?

2. Why should the number of classes in a frequency distribution be between 5 and 20?

3. What is the difference between class limits and class boundaries?

4. What is the difference between relative frequency and cumulative frequency?

5. After constructing an expanded frequency distribution, what should the sum of the relative frequencies be? Explain.

6. What is the difference between a frequency polygon and an ogive?

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. In a frequency distribution, the class width is the distance between the lower and upper limits of a class.

8. The midpoint of a class is the sum of its lower and upper limits divided by two.

9. An ogive is a graph that displays relative frequencies.

10. Class boundaries are used to ensure that consecutive bars of a histogram touch.

In Exercises 11–14, use the given minimum and maximum data entries and the number of classes to find the class width, the lower class limits, and the upper class limits.

11. min = 9, max = 64, 7 classes

12. min = 12, max = 88, 6 classes

13. min = 17, max = 135, 8 classes

14. min = 54, max = 247, 10 classes

Reading a Frequency Distribution *In Exercises 15 and 16, use the given frequency distribution to find the (a) class width, (b) class midpoints, and (c) class boundaries.*

15. **Cleveland, OH High Temperatures (°F)**

Class	Frequency, f
20–30	19
31–41	43
42–52	68
53–63	69
64–74	74
75–85	68
86–96	24

16. **Travel Time to Work (in minutes)**

Class	Frequency, f
0–9	188
10–19	372
20–29	264
30–39	205
40–49	83
50–59	76
60–69	32

17. Use the frequency distribution in Exercise 15 to construct an expanded frequency distribution, as shown in Example 2.

18. Use the frequency distribution in Exercise 16 to construct an expanded frequency distribution, as shown in Example 2.

Graphical Analysis *In Exercises 19 and 20, use the frequency histogram to*

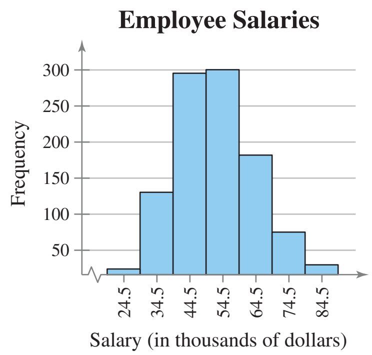
(a) determine the number of classes.

(b) estimate the frequency of the class with the least frequency.

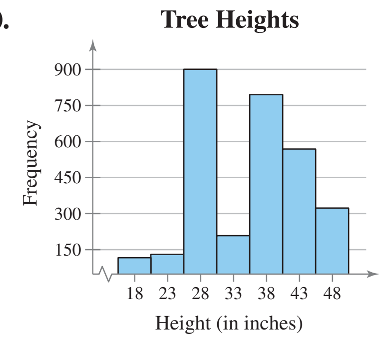
(c) estimate the frequency of the class with the greatest frequency.

(d) determine the class width.

19.



20.

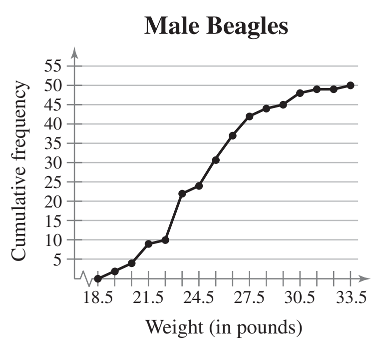


Graphical Analysis *In Exercises 21 and 22, use the ogive to approximate*

(a) the number in the sample.

(b) the location of the greatest increase in frequency.

21.



22.

23. Use the ogive in Exercise 21 to approximate

(a) the cumulative frequency for a weight of 27.5 pounds.

(b) the weight for which the cumulative frequency is 45.

(c) the number of beagles that weigh between 22.5 pounds and 29.5 pounds.

(d) the number of beagles that weigh more than 30.5 pounds.

24. Use the ogive in Exercise 22 to approximate

(a) the cumulative frequency for a height of 72 inches.

(b) the height for which the cumulative frequency is 25.

(c) the number of adult females that are between 62 and 66 inches tall.

(d) the number of adult females that are taller than 70 inches.

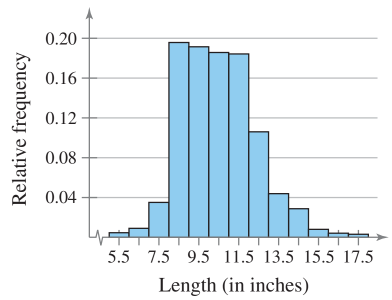
Graphical Analysis *In Exercises 25 and 26, use the relative frequency histogram to*

(a) identify the class with the greatest, and the class with the least, relative frequency.

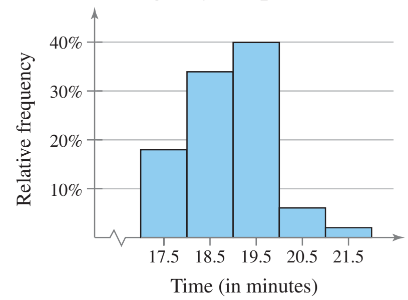
(b) approximate the greatest and least relative frequencies.

(c) approximate the relative frequency of the second class.

25. Atlantic Croaker Fish

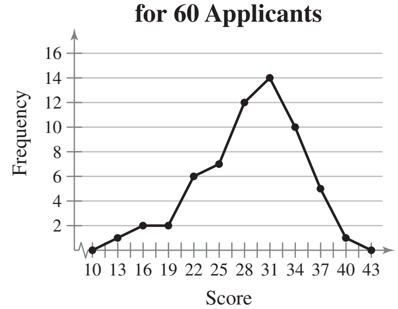


26. Emergency Response Times

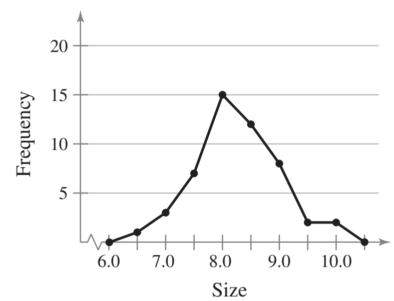


Graphical Analysis *In Exercises 27 and 28, use the frequency polygon to identify the class with the greatest, and the class with the least, frequency.*

27. Raw MCAT Scores for 60 Applicants



28. Shoe Sizes for 50 Females



USING AND INTERPRETING CONCEPTS

Constructing a Frequency Distribution *In Exercises 29 and 30, construct a frequency distribution for the data set using the indicated number of classes. In the table, include the midpoints, relative frequencies, and cumulative frequencies. Which class has the greatest frequency and which has the least frequency?*

29. Political Blog Reading Times
Number of classes: 5
Data set: Time (in minutes) spent reading a political blog in a day

7	39	13	9	25	8	22	0	2	18	2	30	7
35	12	15	8	6	5	29	0	11	39	16	15	

30. Book Spending
Number of classes: 6
Data set: Amount (in dollars) spent on books for a semester

91	472	279	249	530	376	188	341	266	199
142	273	189	130	489	266	248	101	375	486
190	398	188	269	43	30	127	354	84	

indicates that the data set for this exercise is available electronically.

Constructing a Frequency Distribution and a Frequency Histogram

In Exercises 31–34, construct a frequency distribution and a frequency histogram for the data set using the indicated number of classes. Describe any patterns.

31. Sales
Number of classes: 6
Data set: July sales (in dollars) for all sales representatives at a company

2114	2468	7119	1876	4105	3183	1932	1355
4278	1030	2000	1077	5835	1512	1697	2478
3981	1643	1858	1500	4608	1000		

32. Pepper Pungencies
Number of classes: 5
Data set: Pungencies (in 1000s of Scoville units) of 24 tabasco peppers

35	51	44	42	37	38	36	39
44	43	40	40	32	39	41	38
42	39	40	46	37	35	41	39

33. Reaction Times
Number of classes: 8
Data set: Reaction times (in milliseconds) of a sample of 30 adult females to an auditory stimulus

507	389	305	291	336	310	514	442
373	428	387	454	323	441	388	426
411	382	320	450	309	416	359	388
307	337	469	351	422	413		

34. Fracture Times
Number of classes: 5
Data set: Amounts of pressure (in pounds per square inch) at fracture time for 25 samples of brick mortar

2750	2862	2885	2490	2512	2456	2554
2872	2601	2877	2721	2692	2888	2755
2867	2718	2641	2834	2466	2596	2519
2532	2885	2853	2517			

Constructing a Frequency Distribution and a Relative Frequency Histogram

In Exercises 35–38, construct a frequency distribution and a relative frequency histogram for the data set using five classes. Which class has the greatest relative frequency and which has the least relative frequency?

35. Gasoline Consumption
Data set: Highway fuel consumptions (in miles per gallon) for a sample of cars

32	35	28	40	30	42	55	40	45	24
28	34	40	36	34	40	30	25	28	32
40	35	25	44	26	39	38	42	45	32

36. ATM Withdrawals
Data set: A sample of ATM withdrawals (in dollars)

35	10	30	25	75	10	30	20	20	10	40
50	40	30	60	70	25	40	10	60	20	80
40	25	20	10	20	25	30	50	80	20	

37. Triglyceride Levels
Data set: Triglyceride levels (in milligrams per deciliter of blood) of a sample of patients

209	140	155	170	265	138	180	295	250
320	270	225	215	390	420	462	150	200
400	295	240	200	190	145	160	175	

38. Years of Service
Data set: Years of service of a sample of New York state troopers

12	7	9	8	9	8	12	10	9
10	6	8	13	12	10	11	7	14
12	9	8	10	9	11	13	8	

Constructing a Cumulative Frequency Distribution and an Ogive

In Exercises 39 and 40, construct a cumulative frequency distribution and an ogive for the data set using six classes. Then describe the location of the greatest increase in frequency.

39. Retirement Ages
Data set: Retirement ages for a sample of doctors

70	54	55	71	57	58	63	65
60	66	57	62	63	60	63	60
66	60	67	69	69	52	61	73

40. Saturated Fat Intakes
Data set: Daily saturated fat intakes (in grams) of a sample of people

38	32	34	39	40	54	32	17	29	33
57	40	25	36	33	24	42	16	31	33

Constructing a Frequency Distribution and a Frequency Polygon

In Exercises 41 and 42, construct a frequency distribution and a frequency polygon for the data set. Describe any patterns.

41. Exam Scores
Number of classes: 5
Data set: Exam scores for all students in a statistics class

83	92	94	82	73	98	78	85	72	90
89	92	96	89	75	85	63	47	75	82

42. Children of the Presidents
Number of classes: 6
Data set: Number of children of the U.S. presidents
(Source: presidentschildren.com)

0	5	6	0	3	4	0	4	10	15	0	6	2	3	0
4	5	4	8	7	3	5	3	2	6	3	3	1	2	
2	6	1	2	3	2	2	4	4	4	6	1	2	2	

In Exercises 43 and 44, use the data set to construct (a) an expanded frequency distribution, (b) a frequency histogram, (c) a frequency polygon, (d) a relative frequency histogram, and (e) an ogive.

43. Pulse Rates
Number of classes: 6
Data set: Pulse rates of students in a class

68	105	95	80	90	100	75	70	84	98	102	70
65	88	90	75	78	94	110	120	95	80	76	108

44. Hospitals
Number of classes: 8
Data set: Number of hospitals in each state *(Source: American Hospital Directory)*

15	100	56	74	360	53	34	8	213	116
15	38	21	143	97	59	76	110	83	51
23	116	55	91	75	19	108	14	25	14
73	40	30	213	154	97	36	181	12	63
29	121	378	36	91	7	61	71	40	15

SC **45.** Use StatCrunch to construct a frequency histogram and a relative frequency histogram for the following data set that shows the finishing times (in minutes) for 25 runners in a marathon. Use seven classes.

159	164	165	170	215	200	167	225	192	185	235	240	225
191	194	175	167	234	158	172	180	240	176	159	231	

46. Writing What happens when the number of classes is increased for a frequency histogram? Use the data set listed and a technology tool to create frequency histograms with 5, 10, and 20 classes. Which graph displays the data best?

2	7	3	2	11	3	15	8	4	9	10	13	9
7	11	10	1	2	12	5	6	4	2	9	15	

EXTENDING CONCEPTS

47. What Would You Do? You work at a bank and are asked to recommend the amount of cash to put in an ATM each day. You don't want to put in too much (security) or too little (customer irritation). Here are the daily withdrawals (in 100s of dollars) for 30 days.

72	84	61	76	104	76	86	92	80	88
98	76	97	82	84	67	70	81	82	89
74	73	86	81	85	78	82	80	91	83

(a) Construct a relative frequency histogram for the data using 8 classes.

(b) If you put $9000 in the ATM each day, what percent of the days in a month should you expect to run out of cash? Explain your reasoning.

(c) If you are willing to run out of cash for 10% of the days, how much cash should you put in the ATM each day? Explain your reasoning.

48. What Would You Do? You work in the admissions department for a college and are asked to recommend the minimum SAT scores that the college will accept for a position as a full-time student. Here are the SAT scores for a sample of 50 applicants.

1760	1502	1375	1310	1601	1942	1380	2211	1622	1771
1150	1351	1682	1618	2051	1742	1463	1395	1860	1918
1882	1996	1525	1510	2120	1700	1818	1869	1440	1235
976	1513	1790	2250	2102	1905	1979	1588	1420	1730
2175	1930	1965	1658	2005	2125	1260	1560	1635	1620

(a) Construct a relative frequency histogram for the data using 10 classes.

(b) If you set the minimum score at 1616, what percent of the applicants will meet this requirement? Explain your reasoning.

(c) If you want to accept the top 88% of the applicants, what should the minimum score be? Explain your reasoning.

2.2 More Graphs and Displays

WHAT YOU SHOULD LEARN

- How to graph and interpret quantitative data sets using stem-and-leaf plots and dot plots
- How to graph and interpret qualitative data sets using pie charts and Pareto charts
- How to graph and interpret paired data sets using scatter plots and time series charts

STUDY TIP

It is important to include a key for a stem-and-leaf plot to identify the values of the data. This is done by showing a value represented by a stem and one leaf.

Graphing Quantitative Data Sets ▸ Graphing Qualitative Data Sets ▸ Graphing Paired Data Sets

▸ GRAPHING QUANTITATIVE DATA SETS

In Section 2.1, you learned several traditional ways to display quantitative data graphically. In this section, you will learn a newer way to display quantitative data, called a **stem-and-leaf plot.** Stem-and-leaf plots are examples of **exploratory data analysis (EDA),** which was developed by John Tukey in 1977.

In a stem-and-leaf plot, each number is separated into a **stem** (for instance, the entry's leftmost digits) and a **leaf** (for instance, the rightmost digit). You should have as many leaves as there are entries in the original data set and the leaves should be single digits. A stem-and-leaf plot is similar to a histogram but has the advantage that the graph still contains the original data values. Another advantage of a stem-and-leaf plot is that it provides an easy way to sort data.

EXAMPLE 1 SC Report 4

▸ Constructing a Stem-and-Leaf Plot

The following are the numbers of text messages sent last week by the cellular phone users on one floor of a college dormitory. Display the data in a stem-and-leaf plot. What can you conclude?

155	159	144	129	105	145	126	116	130	114	122	112	112	142
126	118	118	108	122	121	109	140	126	119	113	117	118	109
109	119	139	139	122	78	133	126	123	145	121	134	124	119
132	133	124	129	112	126	148	147						

▸ Solution Because the data entries go from a low of 78 to a high of 159, you should use stem values from 7 to 15. To construct the plot, list these stems to the left of a vertical line. For each data entry, list a leaf to the right of its stem. For instance, the entry 155 has a stem of 15 and a leaf of 5. The resulting stem-and-leaf plot will be unordered. To obtain an ordered stem-and-leaf plot, rewrite the plot with the leaves in increasing order from left to right. Be sure to include a key.

Number of Text Messages Sent

7	8	Key: 15│5 = 155
8		
9		
10	5 8 9 9 9	
11	6 4 2 2 8 8 9 3 7 8 9 9 2	
12	9 6 2 6 2 1 6 2 6 3 1 4 4 9 6	
13	0 9 9 3 4 2 3	
14	4 5 2 0 5 8 7	
15	5 9	

Unordered Stem-and-Leaf Plot

Number of Text Messages Sent

7	8	Key: 15│5 = 155
8		
9		
10	5 8 9 9 9	
11	2 2 2 3 4 6 7 8 8 8 9 9 9	
12	1 1 2 2 2 3 4 4 6 6 6 6 6 9 9	
13	0 2 3 3 4 9 9	
14	0 2 4 5 5 7 8	
15	5 9	

Ordered Stem-and-Leaf Plot

Interpretation From the display, you can conclude that more than 50% of the cellular phone users sent between 110 and 130 text messages.

▸ Try It Yourself 1

Use a stem-and-leaf plot to organize the ages of the 50 richest people data set listed in the Chapter Opener on page 37. What can you conclude?

a. List all possible *stems*.
b. List the *leaf* of each data entry to the right of its stem and include a *key*.
c. *Rewrite* the stem-and-leaf plot so that the leaves are ordered.
d. Use the plot to make a *conclusion*. *Answer: Page A31*

EXAMPLE 2

▸ Constructing Variations of Stem-and-Leaf Plots

Organize the data given in Example 1 using a stem-and-leaf plot that has two rows for each stem. What can you conclude?

▸ Solution

Use the stem-and-leaf plot from Example 1, except now list each stem twice. Use the leaves 0, 1, 2, 3, and 4 in the first stem row and the leaves 5, 6, 7, 8, and 9 in the second stem row. The revised stem-and-leaf plot is shown. Notice that by using two rows per stem, you obtain a more detailed picture of the data.

INSIGHT

You can use stem-and-leaf plots to identify unusual data values called *outliers*. In Examples 1 and 2, the data value 78 is an outlier. You will learn more about outliers in Section 2.3.

Number of Text Messages Sent

7	Key: 15$\vert$5 = 155
7	8
8	
8	
9	
9	
10	
10	5 8 9 9 9
11	4 2 2 3 2
11	6 8 8 9 7 8 9 9
12	2 2 1 2 3 1 4 4
12	9 6 6 6 6 9 6
13	0 3 4 2 3
13	9 9
14	4 2 0
14	5 5 8 7
15	
15	5 9

Unordered Stem-and-Leaf Plot

Number of Text Messages Sent

7	Key: 15$\vert$5 = 155
7	8
8	
8	
9	
9	
10	
10	5 8 9 9 9
11	2 2 2 3 4
11	6 7 8 8 8 9 9 9
12	1 1 2 2 2 3 4 4
12	6 6 6 6 6 9 9
13	0 2 3 3 4
13	9 9
14	0 2 4
14	5 5 7 8
15	
15	5 9

Ordered Stem-and-Leaf Plot

Interpretation From the display, you can conclude that most of the cellular phone users sent between 105 and 135 text messages.

▸ Try It Yourself 2

Using two rows for each stem, revise the stem-and-leaf plot you constructed in Try It Yourself 1. What can you conclude?

a. List each stem *twice*.
b. List all leaves *using the appropriate stem row*.
c. Use the plot to make a *conclusion*. *Answer: Page A32*

You can also use a dot plot to graph quantitative data. In a **dot plot,** each data entry is plotted, using a point, above a horizontal axis. Like a stem-and-leaf plot, a dot plot allows you to see how data are distributed, determine specific data entries, and identify unusual data values.

EXAMPLE 3

SC Report 5

Constructing a Dot Plot

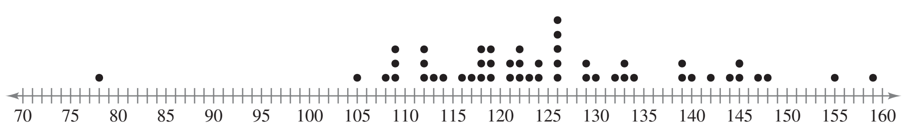
Use a dot plot to organize the text messaging data given in Example 1. What can you conclude from the graph?

155	159	144	129	105	145	126	116	130	114	122	112
112	142	126	118	118	108	122	121	109	140	126	119
113	117	118	109	109	119	139	139	122	78	133	126
123	145	121	134	124	119	132	133	124	129	112	126
148	147										

Solution

So that each data entry is included in the dot plot, the horizontal axis should include numbers between 70 and 160. To represent a data entry, plot a point above the entry's position on the axis. If an entry is repeated, plot another point above the previous point.

Number of Text Messages Sent



Interpretation From the dot plot, you can see that most values cluster between 105 and 148 and the value that occurs the most is 126. You can also see that 78 is an unusual data value.

Try It Yourself 3

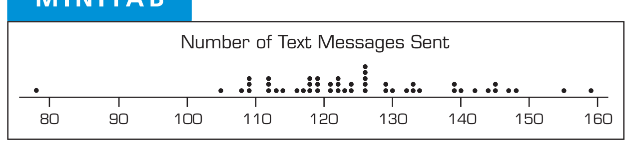
Use a dot plot to organize the ages of the 50 richest people data set listed in the Chapter Opener on page 37. What can you conclude from the graph?

a. Choose an appropriate scale for the *horizontal axis*.
b. Represent each data entry by *plotting a point*.
c. *Describe* any patterns in the data.

Answer: Page A32

Technology can be used to construct stem-and-leaf plots and dot plots. For instance, a MINITAB dot plot for the text messaging data is shown below.

MINITAB



GRAPHING QUALITATIVE DATA SETS

Pie charts provide a convenient way to present qualitative data graphically as percents of a whole. A **pie chart** is a circle that is divided into sectors that represent categories. The area of each sector is proportional to the frequency of each category. In most cases, you will be interpreting a pie chart or constructing one using technology. Example 4 shows how to construct a pie chart by hand.

EXAMPLE 4

SC Report 6

Constructing a Pie Chart

The numbers of earned degrees conferred (in thousands) in 2007 are shown in the table. Use a pie chart to organize the data. What can you conclude? *(Source: U.S. National Center for Education Statistics)*

Earned Degrees Conferred in 2007

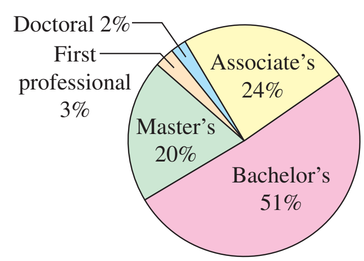
Type of degree	Number (thousands)
Associate's	728
Bachelor's	1525
Master's	604
First professional	90
Doctoral	60

Solution

Begin by finding the relative frequency, or percent, of each category. Then construct the pie chart using the central angle that corresponds to each category. To find the central angle, multiply 360° by the category's relative frequency. For instance, the central angle for associate's degrees is $360°(0.24) \approx 86°$. To construct a pie chart in Excel, follow the instructions in the margin.

Type of degree	f	Relative frequency	Angle
Associate's	728	0.24	86°
Bachelor's	1525	0.51	184°
Master's	604	0.20	72°
First professional	90	0.03	11°
Doctoral	60	0.02	7°

Earned Degrees Conferred in 2007



Interpretation From the pie chart, you can see that over one half of the degrees conferred in 2007 were bachelor's degrees.

Try It Yourself 4

The numbers of earned degrees conferred (in thousands) in 1990 are shown in the table. Use a pie chart to organize the data. Compare the 1990 data with the 2007 data. *(Source: U.S. National Center for Education Statistics)*

Earned Degrees Conferred in 1990

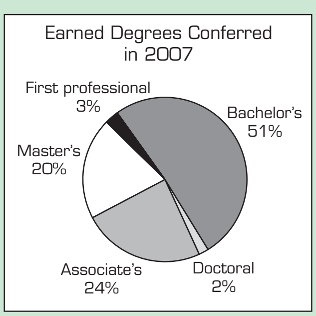
Type of degree	Number (thousands)
Associate's	455
Bachelor's	1052
Master's	325
First professional	71
Doctoral	38

a. Find the *relative frequency* and *central angle* of each category.
b. Use the *central angle* to find the portion that corresponds to each category.
c. *Compare* the 1990 data with the 2007 data.

Answer: Page A32

STUDY TIP

Here are instructions for constructing a pie chart using Excel. First, enter the degree types and their corresponding frequencies or relative frequencies in two separate columns. Then highlight the two columns, click on the Chart Wizard, and select *Pie* as your chart type. Click *Next* throughout the Chart Wizard while constructing your pie chart.



Another way to graph qualitative data is to use a Pareto chart. A **Pareto chart** is a vertical bar graph in which the height of each bar represents frequency or relative frequency. The bars are positioned in order of decreasing height, with the tallest bar positioned at the left. Such positioning helps highlight important data and is used frequently in business.

EXAMPLE 5

SC Report 7

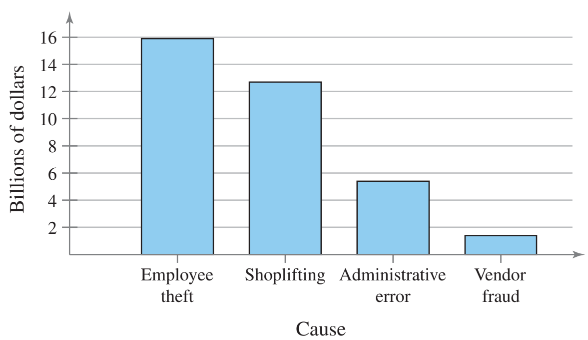
Constructing a Pareto Chart

In a recent year, the retail industry lost $36.5 billion in inventory shrinkage. Inventory shrinkage is the loss of inventory through breakage, pilferage, shoplifting, and so on. The main causes of inventory shrinkage are administrative error ($5.4 billion), employee theft ($15.9 billion), shoplifting ($12.7 billion), and vendor fraud ($1.4 billion). If you were a retailer, which causes of inventory shrinkage would you address first? *(Source: National Retail Federation and the University of Florida)*

Solution

Using frequencies for the vertical axis, you can construct the Pareto chart as shown.

Main Causes of Inventory Shrinkage



Interpretation From the graph, it is easy to see that the causes of inventory shrinkage that should be addressed first are employee theft and shoplifting.

Try It Yourself 5

Every year, the Better Business Bureau (BBB) receives complaints from customers. In a recent year, the BBB received the following complaints.

7792 complaints about home furnishing stores
5733 complaints about computer sales and service stores
14,668 complaints about auto dealers
9728 complaints about auto repair shops
4649 complaints about dry cleaning companies

Use a Pareto chart to organize the data. What source is the greatest cause of complaints? *(Source: Council of Better Business Bureaus)*

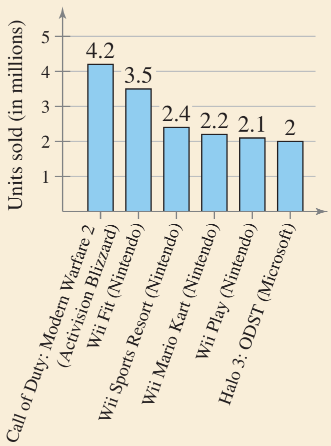
a. Find the *frequency or relative frequency* for each data entry.
b. *Position the bars in decreasing order* according to frequency or relative frequency.
c. *Interpret* the results in the context of the data.

Answer: Page A32

PICTURING THE WORLD

The six top-selling video games of 2009 through November are shown in the following Pareto chart. Publishers are in parentheses. (Source: NPD Group)

Six Top-Selling Video Games of 2009 Through November



Of the six top-selling video games, how many units did Nintendo sell?

▸ GRAPHING PAIRED DATA SETS

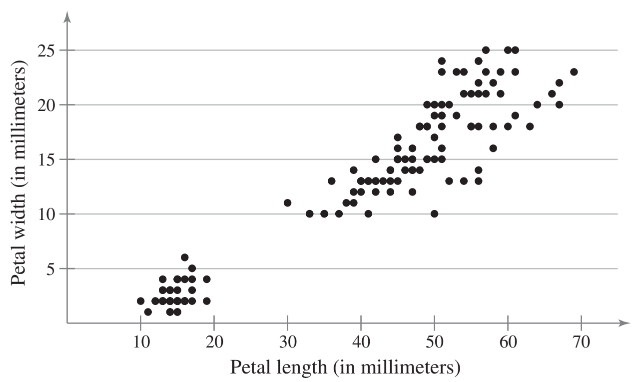
When each entry in one data set corresponds to one entry in a second data set, the sets are called **paired data sets.** For instance, suppose a data set contains the costs of an item and a second data set contains sales amounts for the item at each cost. Because each cost corresponds to a sales amount, the data sets are paired. One way to graph paired data sets is to use a **scatter plot,** where the ordered pairs are graphed as points in a coordinate plane. A scatter plot is used to show the relationship between two quantitative variables.

EXAMPLE 6

▸ Interpreting a Scatter Plot

The British statistician Ronald Fisher (see page 33) introduced a famous data set called Fisher's Iris data set. This data set describes various physical characteristics, such as petal length and petal width (in millimeters), for three species of iris. In the scatter plot shown, the petal lengths form the first data set and the petal widths form the second data set. As the petal length increases, what tends to happen to the petal width? *(Source: Fisher, R. A., 1936)*

Fisher's Iris Data Set



▸ Solution

The horizontal axis represents the petal length, and the vertical axis represents the petal width. Each point in the scatter plot represents the petal length and petal width of one flower.

Interpretation From the scatter plot, you can see that as the petal length increases, the petal width also tends to increase.

▸ Try It Yourself 6

The lengths of employment and the salaries of 10 employees are listed in the table at the left. Graph the data using a scatter plot. What can you conclude?

Length of employment (in years)	Salary (in dollars)
5	32,000
4	32,500
8	40,000
4	27,350
2	25,000
10	43,000
7	41,650
6	39,225
9	45,100
3	28,000

a. Label the *horizontal and vertical axes*.
b. *Plot* the paired data.
c. *Describe* any trends.

Answer: Page A32

You will learn more about scatter plots and how to analyze them in Chapter 9.

A data set that is composed of quantitative entries taken at regular intervals over a period of time is called a **time series.** For instance, the amount of precipitation measured each day for one month is a time series. You can use a **time series chart** to graph a time series.

EXAMPLE 7

SC Report 8

See MINITAB and TI-83/84 Plus steps on pages 122 and 123.

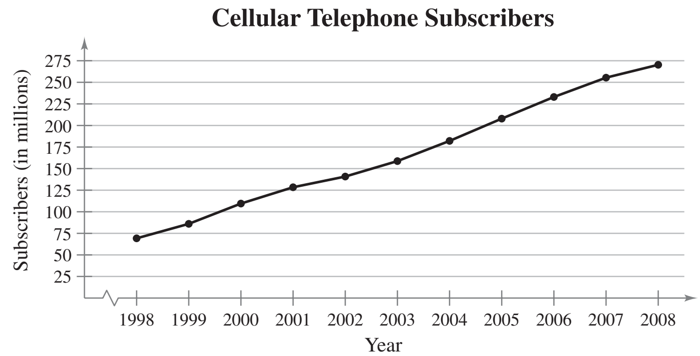
Constructing a Time Series Chart

The table lists the number of cellular telephone subscribers (in millions) and subscribers' average local monthly bills for service (in dollars) for the years 1998 through 2008. Construct a time series chart for the number of cellular subscribers. What can you conclude? *(Source: Cellular Telecommunications & Internet Association)*

Year	Subscribers (in millions)	Average bill (in dollars)
1998	69.2	39.43
1999	86.0	41.24
2000	109.5	45.27
2001	128.4	47.37
2002	140.8	48.40
2003	158.7	49.91
2004	182.1	50.64
2005	207.9	49.98
2006	233.0	50.56
2007	255.4	49.79
2008	270.3	50.07

Solution

Let the horizontal axis represent the years and let the vertical axis represent the number of subscribers (in millions). Then plot the paired data and connect them with line segments.



Interpretation The graph shows that the number of subscribers has been increasing since 1998.

Try It Yourself 7

Use the table in Example 7 to construct a time series chart for subscribers' average local monthly cellular telephone bills for the years 1998 through 2008. What can you conclude?

a. Label the *horizontal and vertical axes*.
b. *Plot* the paired data and *connect* them with line segments.
c. *Describe* any patterns you see.

Answer: Page A32

2.2 EXERCISES

BUILDING BASIC SKILLS AND VOCABULARY

1. Name some ways to display quantitative data graphically. Name some ways to display qualitative data graphically.

2. What is an advantage of using a stem-and-leaf plot instead of a histogram? What is a disadvantage?

3. In terms of displaying data, how is a stem-and-leaf plot similar to a dot plot?

4. How is a Pareto chart different from a standard vertical bar graph?

Putting Graphs in Context *In Exercises 5–8, match the plot with the description of the sample.*

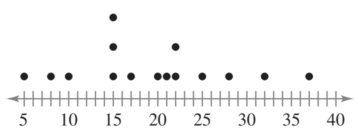
5. Key: 0|8 = 0.8

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

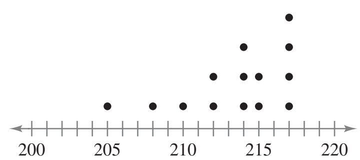
6. Key: 6|7 = 67

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

7.



8.



(a) Time (in minutes) it takes a sample of employees to drive to work

(b) Grade point averages of a sample of students with finance majors

(c) Top speeds (in miles per hour) of a sample of high-performance sports cars

(d) Ages (in years) of a sample of residents of a retirement home

Graphical Analysis *In Exercises 9–12, use the stem-and-leaf plot or dot plot to list the actual data entries. What is the maximum data entry? What is the minimum data entry?*

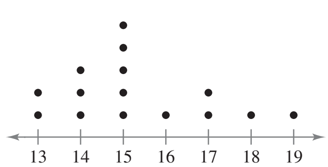
9. Key: 2|7 = 27

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

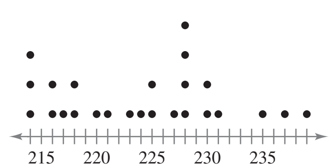
10. Key: 12|9 = 12.9

12	
12	9
13	3
13	6 7 7
14	1 1 1 1 3 4 4
14	6 9 9
15	0 0 0 1 2 4
15	6 7 8 8 8 9
16	1
16	6 7

11.



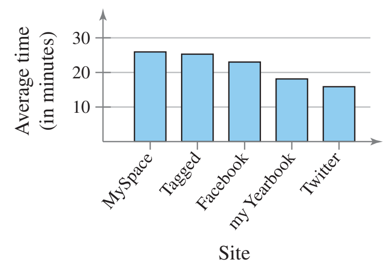
12.



USING AND INTERPRETING CONCEPTS

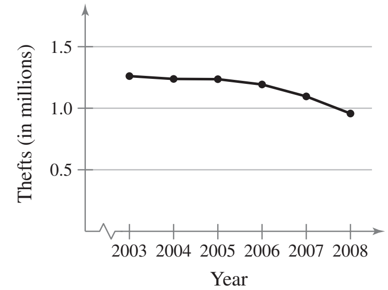
Graphical Analysis *In Exercises 13–16, give three conclusions that can be drawn from the graph.*

13. **Average Time Spent on Top 5 Social Networking Sites**



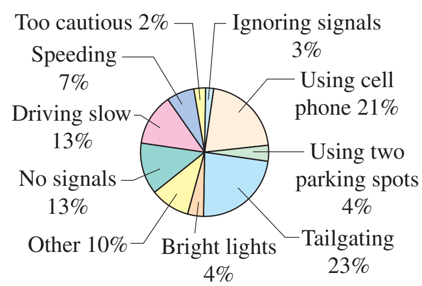
(Source: Experian Hitwise)

14. **Motor Vehicle Thefts in U.S.**



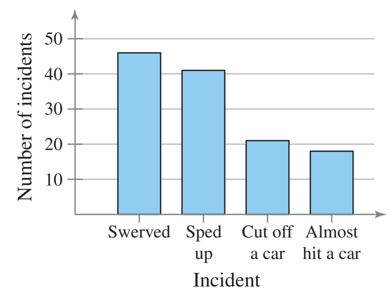
(Source: Federal Bureau of Investigation)

15. **How Other Drivers Irk Us**



(Adapted from Reuters/Zogby)

16. **Driving and Cell Phone Use**



(Adapted from USA Today)

Graphing Data Sets *In Exercises 17–30, organize the data using the indicated type of graph. What can you conclude about the data?*

17. Exam Scores Use a stem-and-leaf plot to display the data. The data represent the scores of a biology class on a midterm exam.

75	85	90	80	87	67	82	88	95	91	73	80
83	92	94	68	75	91	79	95	87	76	91	85

18. Highest Paid CEOs Use a stem-and-leaf plot that has two rows for each stem to display the data. The data represent the ages of the top 30 highest paid CEOs. *(Source: Forbes)*

64	74	55	55	62	63	50	67	51	59	50
52	50	59	62	64	57	61	49	63	62	60
55	56	48	58	64	60	60	57			

19. Ice Thickness Use a stem-and-leaf plot to display the data. The data represent the thicknesses (in centimeters) of ice measured at 20 different locations on a frozen lake.

5.8	6.4	6.9	7.2	5.1	4.9	4.3	5.8	7.0	6.8
8.1	7.5	7.2	6.9	5.8	7.2	8.0	7.0	6.9	5.9

20. Apple Prices Use a stem-and-leaf plot to display the data. The data represent the prices (in cents per pound) paid to 28 farmers for apples.

19.2	19.6	16.4	17.1	19.0	17.4	17.3
20.1	19.0	17.5	17.6	18.6	18.4	17.7
19.5	18.4	18.9	17.5	19.3	20.8	19.3
18.6	18.6	18.3	17.1	18.1	16.8	17.9

21. Systolic Blood Pressures Use a dot plot to display the data. The data represent the systolic blood pressures (in millimeters of mercury) of 30 patients at a doctor's office.

120	135	140	145	130	150	120	170	145	125
130	110	160	180	200	150	200	135	140	120
120	130	140	170	120	165	150	130	135	140

22. Life Spans of Houseflies Use a dot plot to display the data. The data represent the life spans (in days) of 40 houseflies.

9	9	4	4	8	11	10	5	8	13	9
6	7	11	13	11	6	9	8	14	10	6
10	10	8	7	14	11	7	8	6	11	13
10	14	14	8	13	14	10				

23. New York City Marathon Use a pie chart to display the data. The data represent the number of men's New York City Marathon winners from each country through 2009. *(Source: New York Road Runners)*

United States	15	Mexico	4
Italy	4	Morocco	1
Ethiopia	1	Great Britain	1
South Africa	2	Brazil	2
Tanzania	1	New Zealand	1
Kenya	8		

24. NASA Budget Use a pie chart to display the data. The data represent the 2010 NASA budget request (in millions of dollars) divided among five categories. *(Source: NASA)*

Science, aeronautics, exploration	8947
Space operations	6176
Education	126
Cross-agency support	3401
Inspector general	36

25. Barrel of Oil Use a Pareto chart to display the data. The data represent how a 42-gallon barrel of crude oil is distributed. *(Adapted from American Petroleum Institute)*

Gasoline	43%
Kerosene-type jet fuel	9%
Distillate fuel oil (home heating, diesel fuel, etc.)	24%
Coke	5%
Residual fuel oil (industry, marine transportation, etc.)	4%
Liquefied refinery gases	3%
Other	12%

26. UV Index Use a Pareto chart to display the data. The data represent the ultraviolet indices for five cities at noon on a recent date. *(Source: National Oceanic and Atmospheric Administration)*

Atlanta, GA	Boise, ID	Concord, NH	Denver, CO	Miami, FL
9	7	8	7	10

27. Hourly Wages Use a scatter plot to display the data shown in the table. The data represent the number of hours worked and the hourly wages (in dollars) for a sample of 12 production workers. Describe any trends shown.

Hours	Hourly wage
33	12.16
37	9.98
34	10.79
40	11.71
35	11.80
33	11.51
40	13.65
33	12.05
28	10.54
45	10.33
37	11.57
28	10.17

TABLE FOR EXERCISE 27

Number of students per teacher	Average teacher's salary
17.1	28.7
17.5	47.5
18.9	31.8
17.1	28.1
20.0	40.3
18.6	33.8
14.4	49.8
16.5	37.5
13.3	42.5
18.4	31.9

TABLE FOR EXERCISE 28

28. Salaries Use a scatter plot to display the data shown in the table. The data represent the number of students per teacher and the average teacher salaries (in thousands of dollars) for a sample of 10 school districts. Describe any trends shown.

29. Daily High Temperatures Use a time series chart to display the data. The data represent the daily high temperatures for a city for a period of 12 days.

May 1	May 2	May 3	May 4	May 5	May 6
77°	77°	79°	81°	82°	82°

May 7	May 8	May 9	May 10	May 11	May 12
85°	87°	90°	88°	89°	82°

30. Manufacturing Use a time series chart to display the data. The data represent the percentages of the U.S. gross domestic product (GDP) that come from the manufacturing sector. *(Source: U.S. Bureau of Economic Analysis)*

1997	1998	1999	2000	2001	2002
16.6%	15.4%	14.8%	14.5%	13.2%	12.9%

2003	2004	2005	2006	2007	2008
12.5%	12.2%	11.9%	12.0%	11.7%	11.5%

SC *In Exercises 31–34, use StatCrunch to organize the data using the indicated type of graph. What can you conclude about the data?*

31. Use a stem-and-leaf plot to display the data. The data represent the scores of an economics class on a final exam.

82 93 95 75 68 90 98 71 85 88 100 93
70 80 89 62 55 95 83 86 88 76 99 87

32. Use a dot plot to display the data. The data represent the screen sizes (in inches) of 20 DVD camcorders.

3.0 2.7 3.2 2.7 1.8 2.7 2.7 3.0 2.7 3.0
2.5 3.2 2.7 2.7 3.0 2.7 2.0 2.7 3.0 2.5

33. Use (a) a pie chart and (b) a Pareto chart to display the data. The data represent the results of an online survey that asked adults which type of investment they would focus on in 2010. *(Adapted from CNN)*

U.S. stocks	11,521	Emerging markets	5267
Bonds	3292	Commodities	1975
Bank accounts	10,533		

34. The data represent the number of motor vehicles (in millions) registered in the U.S. and the number of crashes (in millions). *(Source: U.S. National Highway Safety Traffic Administration)*

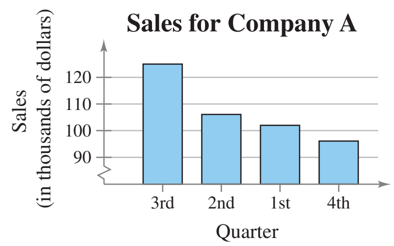
Year	2000	2001	2002	2003	2004	2005	2006	2007
Registrations	221	230	230	231	237	241	244	247
Crashes	6.4	6.3	6.3	6.3	6.2	6.2	6.0	6.0

(a) Use a scatter plot to display the number of registrations.
(b) Use a scatter plot to display the number of crashes.
(c) Construct a time series chart for the number of registrations.
(d) Construct a time series chart for the number of crashes.

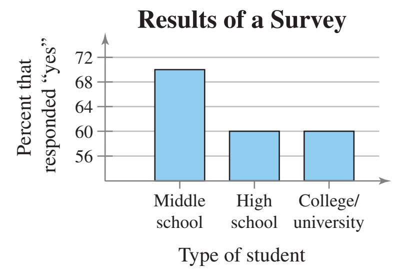
EXTENDING CONCEPTS

A Misleading Graph? *A misleading graph is a statistical graph that is not drawn appropriately. This type of graph can misrepresent data and lead to false conclusions. In Exercises 35–38, (a) explain why the graph is misleading, and (b) redraw the graph so that it is not misleading.*

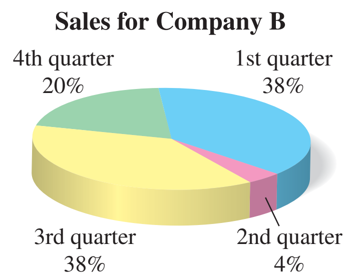
35.



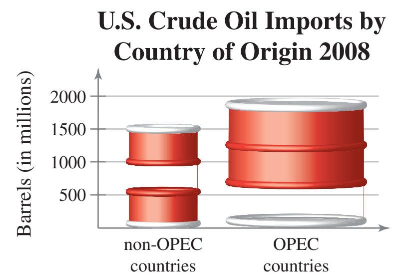
36.



37.



38.



39. Law Firm Salaries A **back-to-back stem-and-leaf plot** compares two data sets by using the same stems for each data set. Leaves for the first data set are on one side while leaves for the second data set are on the other side. The back-to-back stem-and-leaf plot shows the salaries (in thousands of dollars) of all lawyers at two small law firms.

(a) What are the lowest and highest salaries at Law Firm A? at Law Firm B?

(b) How many lawyers are in each firm?

(c) Compare the distribution of salaries at each law firm. What do you notice?

Law Firm A		Law Firm B
5 0	9	0 3
8 5 2 2 2	10	5 7
9 9 7 0 0	11	0 0 5
1 1	12	0 3 3 5
	13	2 2 5 9
	14	1 3 3 3 9
	15	5 5 5 6
	16	4 9 9
9 9 5 1 0	17	1 2 5
5 5 5 2 1	18	9
9 9 8 7 5	19	0
3	20	

Key: 5|19|0 = $195,000 for Law Firm A and $190,000 for Law Firm B

FIGURE FOR EXERCISE 39

40. Yoga Classes The data sets show the ages of all participants in two yoga classes.

3:00 P.M. Class

40	60	73	77	51	68
68	35	68	53	64	75
76	69	59	55	38	57
68	84	75	62	73	75
85	77				

8:00 P.M. Class

19	18	20	29	39	43
71	56	44	44	18	19
19	18	18	20	25	29
25	22	31	24	24	23
19	19	18	28	20	31

(a) Make a back-to-back stem-and-leaf plot to display the data.

(b) What are the lowest and highest ages of participants in the 3:00 P.M. class? in the 8:00 P.M. class?

(c) How many participants are in each class?

(d) Compare the distribution of ages in each class. What conclusion(s) can you make based on your observations?

2.3 Measures of Central Tendency

WHAT YOU SHOULD LEARN

- How to find the mean, median, and mode of a population and of a sample
- How to find a weighted mean of a data set and the mean of a frequency distribution
- How to describe the shape of a distribution as symmetric, uniform, or skewed and how to compare the mean and median for each

Mean, Median, and Mode ▸ Weighted Mean and Mean of Grouped Data ▸ The Shapes of Distributions

MEAN, MEDIAN, AND MODE

In Sections 2.1 and 2.2, you learned about the graphical representations of quantitative data. In Sections 2.3 and 2.4, you will learn how to supplement graphical representations with numerical statistics that describe the center and variability of a data set.

A **measure of central tendency** is a value that represents a typical, or central, entry of a data set. The three most commonly used measures of central tendency are the *mean*, the *median*, and the *mode*.

DEFINITION

The **mean** of a data set is the sum of the data entries divided by the number of entries. To find the mean of a data set, use one of the following formulas.

Population Mean: $\mu = \dfrac{\Sigma x}{N}$ Sample Mean: $\bar{x} = \dfrac{\Sigma x}{n}$

The lowercase Greek letter μ (pronounced mu) represents the population mean and $\bar{x}$ (read as "x bar") represents the sample mean. Note that N represents the number of entries in a *population* and n represents the number of entries in a *sample*. Recall that the uppercase Greek letter sigma (Σ) indicates a summation of values.

EXAMPLE 1

SC Report 9

Finding a Sample Mean

The prices (in dollars) for a sample of round-trip flights from Chicago, Illinois to Cancun, Mexico are listed. What is the mean price of the flights?

872 432 397 427 388 782 397

Solution

The sum of the flight prices is

$$\Sigma x = 872 + 432 + 397 + 427 + 388 + 782 + 397 = 3695.$$

To find the mean price, divide the sum of the prices by the number of prices in the sample.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{3695}{7} \approx 527.9$$

So, the mean price of the flights is about \$527.90.

STUDY TIP

Notice that the mean in Example 1 has one more decimal place than the original set of data values. This *round-off rule* will be used throughout the text. Another important *round-off rule* is that rounding should not be done until the final answer of a calculation.

Try It Yourself 1

The heights (in inches) of the players on the 2009–2010 Cleveland Cavaliers basketball team are listed. What is the mean height?

Heights of Players							
74	78	81	87	81	80	77	80
85	78	80	83	75	81	73	

a. *Find the sum* of the data entries.
b. *Divide the sum* by the number of data entries.
c. *Interpret* the results in the context of the data.

Answer: Page A32

DEFINITION

The **median** of a data set is the value that lies in the middle of the data when the data set is ordered. The median measures the center of an ordered data set by dividing it into two equal parts. If the data set has an odd number of entries, the median is the middle data entry. If the data set has an even number of entries, the median is the mean of the two middle data entries.

EXAMPLE 2

SC Report 10

Finding the Median

Find the median of the flight prices given in Example 1.

Solution

To find the median price, first order the data.

388 397 397 427 432 782 872

Because there are seven entries (an odd number), the median is the middle, or fourth, data entry. So, the median flight price is \$427.

Try It Yourself 2

The ages of a sample of fans at a rock concert are listed. Find the median age.

24 27 19 21 18 23 21 20 19 33 30 29 21
18 24 26 38 19 35 34 33 30 21 27 30

a. *Order* the data entries.
b. *Find the middle* data entry.
c. *Interpret* the results in the context of the data.

Answer: Page A32

STUDY TIP

In a data set, there are the same number of data values above the median as there are below the median. For instance, in Example 2, three of the prices are below \$427 and three are above \$427.

EXAMPLE 3

Finding the Median

In Example 2, the flight priced at \$432 is no longer available. What is the median price of the remaining flights?

Solution

The remaining prices, in order, are 388, 397, 397, 427, 782, and 872.

Because there are six entries (an even number), the median is the mean of the two middle entries.

$$\text{Median} = \frac{397 + 427}{2} = 412$$

So, the median price of the remaining flights is \$412.

Try It Yourself 3

The prices (in dollars) of a sample of digital photo frames are listed. Find the median price of the digital photo frames.

25 100 130 60 140 200 220 80 250 97

a. *Order* the data entries.
b. *Find the mean* of the two middle data entries.
c. *Interpret* the results in the context of the data.

Answer: Page A32

DEFINITION

The **mode** of a data set is the data entry that occurs with the greatest frequency. A data set can have one mode, more than one mode, or no mode. If no entry is repeated, the data set has no mode. If two entries occur with the same greatest frequency, each entry is a mode and the data set is called **bimodal.**

INSIGHT

The mode is the only measure of central tendency that can be used to describe data at the nominal level of measurement. But when working with quantitative data, the mode is rarely used.

EXAMPLE 4

SC Report 11

Finding the Mode

Find the mode of the flight prices given in Example 1.

Solution

Ordering the data helps to find the mode.

388 397 397 427 432 782 872

From the ordered data, you can see that the entry 397 occurs twice, whereas the other data entries occur only once. So, the mode of the flight prices is \$397.

Try It Yourself 4

The prices (in dollars per square foot) for a sample of South Beach (Miami Beach, FL) condominiums are listed. Find the mode of the prices.

324	462	540	450	638	564	670	618	624	825
540	980	1650	1420	670	830	912	750	1260	450
975	670	1100	980	750	723	705	385	475	720

a. Write the data in *order*.
b. Identify the entry, or entries, that occur with the *greatest frequency*.
c. *Interpret* the results in the context of the data. *Answer: Page A32*

EXAMPLE 5

Political party	Frequency, f
Democrat	34
Republican	56
Other	21
Did not respond	9

Finding the Mode

At a political debate, a sample of audience members were asked to name the political party to which they belonged. Their responses are shown in the table. What is the mode of the responses?

Solution

The response occurring with the greatest frequency is Republican. So, the mode is Republican.

Interpretation In this sample, there were more Republicans than people of any other single affiliation.

Try It Yourself 5

In a survey, 1000 U.S. adults were asked if they thought public cellular phone conversations were rude. Of those surveyed, 510 responded "Yes," 370 responded "No," and 120 responded "Not sure." What is the mode of the responses? *(Adapted from Fox TV/Rasmussen Reports)*

a. Identify the entry that occurs with the *greatest frequency*.
b. *Interpret* the results in the context of the data. *Answer: Page A32*

Although the mean, the median, and the mode each describe a typical entry of a data set, there are advantages and disadvantages of using each. The mean is a reliable measure because it takes into account every entry of a data set. However, the mean can be greatly affected when the data set contains *outliers*.

DEFINITION

An **outlier** is a data entry that is far removed from the other entries in the data set.

Ages in a Class						
20	20	20	20	20	20	21
21	21	21	22	22	22	23
23	23	23	24	24	65	

Outlier → 65

A data set can have one or more outliers, causing **gaps** in a distribution. Conclusions that are drawn from a data set that contains outliers may be flawed.

EXAMPLE 6

Comparing the Mean, the Median, and the Mode

Find the mean, the median, and the mode of the sample ages of students in a class shown at the left. Which measure of central tendency best describes a typical entry of this data set? Are there any outliers?

Solution

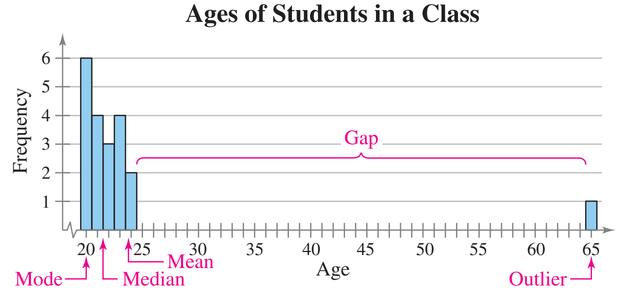
Mean: $\bar{x} = \frac{\Sigma x}{n} = \frac{475}{20} \approx 23.8$ years

Median: Median $= \frac{21 + 22}{2} = 21.5$ years

Mode: The entry occurring with the greatest frequency is 20 years.

Interpretation The mean takes every entry into account but is influenced by the outlier of 65. The median also takes every entry into account, and it is not affected by the outlier. In this case the mode exists, but it doesn't appear to represent a typical entry. Sometimes a graphical comparison can help you decide which measure of central tendency best represents a data set. The histogram shows the distribution of the data and the locations of the mean, the median, and the mode. In this case, it appears that the median best describes the data set.

Ages of Students in a Class



Try It Yourself 6

Remove the data entry 65 from the data set in Example 6. Then rework the example. How does the absence of this outlier change each of the measures?

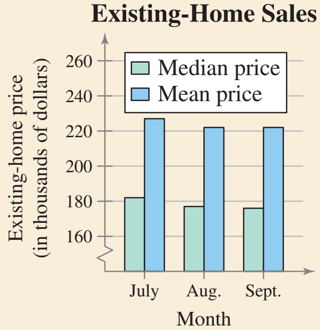
a. Find the *mean*, the *median*, and the *mode*.
b. *Compare* these measures of central tendency with those found in Example 6.

Answer: Page A33

PICTURING THE WORLD

The National Association of Realtors keeps a databank of existing-home sales. One list uses the *median* price of existing homes sold and another uses the *mean* price of existing homes sold. The sales for the third quarter of 2009 are shown in the double-bar graph.
(Source: National Association of Realtors)

2009 U.S. Existing-Home Sales



Notice in the graph that each month the mean price is about $45,000 more than the median price. What factors would cause the mean price to be greater than the median price?

WEIGHTED MEAN AND MEAN OF GROUPED DATA

Sometimes data sets contain entries that have a greater effect on the mean than do other entries. To find the mean of such a data set, you must find the *weighted mean*.

DEFINITION

A **weighted mean** is the mean of a data set whose entries have varying weights. A weighted mean is given by

$$\bar{x} = \frac{\sum (x \cdot w)}{\sum w}$$

where w is the weight of each entry x.

EXAMPLE 7

Finding a Weighted Mean

You are taking a class in which your grade is determined from five sources: 50% from your test mean, 15% from your midterm, 20% from your final exam, 10% from your computer lab work, and 5% from your homework. Your scores are 86 (test mean), 96 (midterm), 82 (final exam), 98 (computer lab), and 100 (homework). What is the weighted mean of your scores? If the minimum average for an A is 90, did you get an A?

Solution

Begin by organizing the scores and the weights in a table.

Source	Score, x	Weight, w	xw
Test mean	86	0.50	43.0
Midterm	96	0.15	14.4
Final exam	82	0.20	16.4
Computer lab	98	0.10	9.8
Homework	100	0.05	5.0
		$\sum w = 1$	$\sum (x \cdot w) = 88.6$

$$\bar{x} = \frac{\sum (x \cdot w)}{\sum w}$$

$$= \frac{88.6}{1}$$

$$= 88.6$$

Your weighted mean for the course is 88.6. So, you did not get an A.

Try It Yourself 7

An error was made in grading your final exam. Instead of getting 82, you scored 98. What is your new weighted mean?

a. Multiply each score by its weight and *find the sum of these products*.
b. Find the *sum of the weights*.
c. Find the *weighted mean*.
d. *Interpret* the results in the context of the data.

Answer: Page A33

If data are presented in a frequency distribution, you can approximate the mean as follows.

STUDY TIP

If the frequency distribution represents a population, then the mean of the frequency distribution is approximated by

$$\mu = \frac{\Sigma(x \cdot f)}{N}$$

where $N = \Sigma f$.

DEFINITION

The **mean of a frequency distribution** for a sample is approximated by

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} \qquad \text{Note that } n = \Sigma f.$$

where x and f are the midpoints and frequencies of a class, respectively.

GUIDELINES

Finding the Mean of a Frequency Distribution

IN WORDS	IN SYMBOLS
1. Find the midpoint of each class.	$x = \frac{(\text{Lower limit}) + (\text{Upper limit})}{2}$
2. Find the sum of the products of the midpoints and the frequencies.	$\Sigma(x \cdot f)$
3. Find the sum of the frequencies.	$n = \Sigma f$ inconsistence
4. Find the mean of the frequency distribution.	$\bar{x} = \frac{\Sigma(x \cdot f)}{n}$

EXAMPLE 8

Finding the Mean of a Frequency Distribution

Use the frequency distribution at the left to approximate the mean number of minutes that a sample of Internet subscribers spent online during their most recent session.

Class midpoint, x	Frequency, f	xf
12.5	6	75.0
24.5	10	245.0
36.5	13	474.5
48.5	8	388.0
60.5	5	302.5
72.5	6	435.0
84.5	2	169.0
	$n = 50$	$\Sigma = 2089.0$

Solution

$$\bar{x} = \frac{\Sigma(x \cdot f)}{n} = \frac{2089.0}{50} \approx 41.8$$

So, the mean time spent online was approximately 41.8 minutes.

Try It Yourself 8

Use a frequency distribution to approximate the mean age of the 50 richest people. (See Try It Yourself 2 on page 41.)

a. Find the *midpoint* of each class.
b. Find the *sum of the products* of each midpoint and corresponding frequency.
c. Find the *sum of the frequencies*.
d. Find the *mean of the frequency distribution*.

Answer: Page A33

THE SHAPES OF DISTRIBUTIONS

A graph reveals several characteristics of a frequency distribution. One such characteristic is the shape of the distribution.

DEFINITION

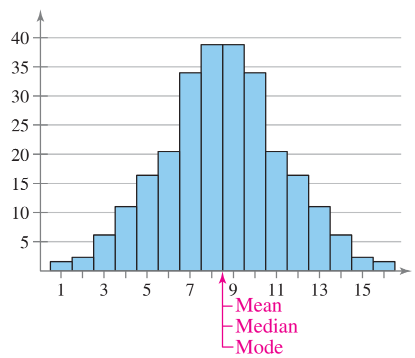
A frequency distribution is **symmetric** when a vertical line can be drawn through the middle of a graph of the distribution and the resulting halves are approximately mirror images.

A frequency distribution is **uniform** (or **rectangular**) when all entries, or classes, in the distribution have equal or approximately equal frequencies. A uniform distribution is also symmetric.

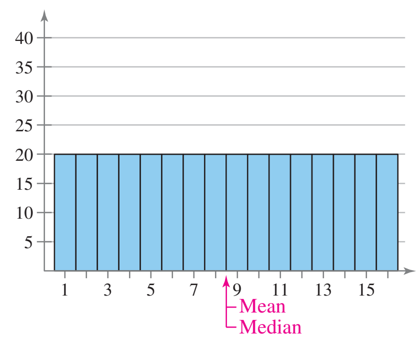
A frequency distribution is skewed if the "tail" of the graph elongates more to one side than to the other. A distribution is **skewed left (negatively skewed)** if its tail extends to the left. A distribution is **skewed right (positively skewed)** if its tail extends to the right.

To explore this topic further, see Activity 2.3 on page 79.

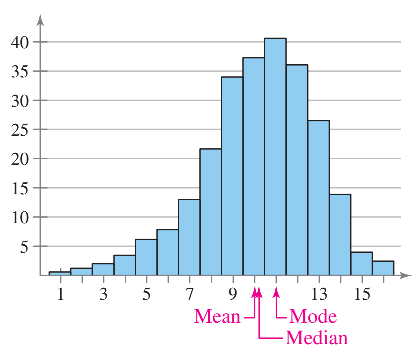
When a distribution is symmetric and unimodal, the mean, median, and mode are equal. If a distribution is skewed left, the mean is less than the median and the median is usually less than the mode. If a distribution is skewed right, the mean is greater than the median and the median is usually greater than the mode. Examples of these commonly occurring distributions are shown.



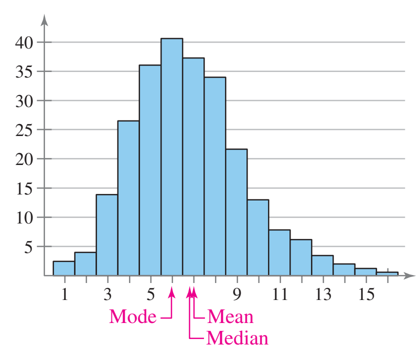
Symmetric Distribution



Uniform Distribution



Skewed Left Distribution



Skewed Right Distribution

INSIGHT

Be aware that there are many different shapes of distributions. In some cases, the shape cannot be classified as symmetric, uniform, or skewed. A distribution can have several gaps caused by outliers, or **clusters** of data. Clusters occur when several types of data are included in the one data set.

The mean will always fall in the direction in which the distribution is skewed. For instance, when a distribution is skewed left, the mean is to the left of the median.

2.3 EXERCISES

BUILDING BASIC SKILLS AND VOCABULARY

True or False? *In Exercises 1–4, determine whether the statement is true or false. If it is false, rewrite it as a true statement.*

1. The mean is the measure of central tendency most likely to be affected by an outlier.

2. Some quantitative data sets do not have medians.

3. A data set can have the same mean, median, and mode.

4. When each data class has the same frequency, the distribution is symmetric.

Constructing Data Sets *In Exercises 5–8, construct the described data set. The values in the data set cannot all be the same.*

5. Median and mode are the same.

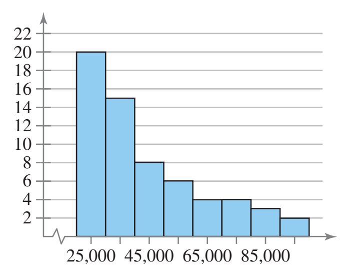
6. Mean and mode are the same.

7. Mean is *not* representative of a typical number in the data set.

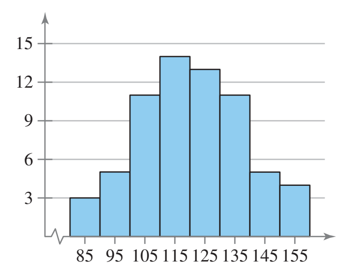
8. Mean, median, and mode are the same.

Graphical Analysis *In Exercises 9–12, determine whether the approximate shape of the distribution in the histogram is symmetric, uniform, skewed left, skewed right, or none of these. Justify your answer.*

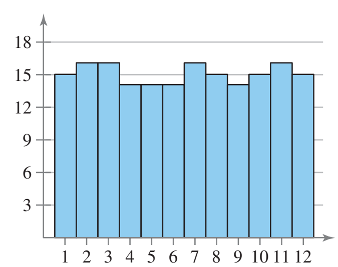
9.



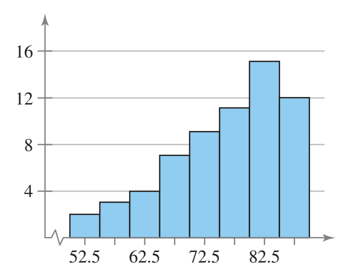
10.



11.



12.



Matching *In Exercises 13–16, match the distribution with one of the graphs in Exercises 9–12. Justify your decision.*

13. The frequency distribution of 180 rolls of a dodecagon (a 12-sided die)

14. The frequency distribution of salaries at a company where a few executives make much higher salaries than the majority of employees

15. The frequency distribution of scores on a 90-point test where a few students scored much lower than the majority of students

16. The frequency distribution of weights for a sample of seventh grade boys

USING AND INTERPRETING CONCEPTS

Finding and Discussing the Mean, Median, and Mode *In Exercises 17–34, find the mean, median, and mode of the data, if possible. If any of these measures cannot be found or a measure does not represent the center of the data, explain why.*

17. Concert Tickets The number of concert tickets purchased online for the last 13 purchases

4 2 5 8 6 6 4 3 2 4 7 8 5

18. Tuition The 2009–2010 tuition and fees (in thousands of dollars) for the top 10 liberal arts colleges *(Source: U.S. News and World Report)*

39 39 38 51 38 40 37 40 35 39

19. MCAT Scores The average medical college admission test (MCAT) scores for a sample of seven medical schools *(Source: Association of American Medical Colleges)*

11.0 11.7 10.3 11.7 11.7 10.7 9.7

20. Cholesterol The cholesterol levels of a sample of 10 female employees

154 240 171 188 235 203 184 173 181 275

21. NFL The average points per game scored by each NFL team during the 2009 regular season *(Source: National Football League)*

20.4 19.7 17.5 26.7 22.7 21.8 16.6 29.4
26.0 22.5 28.8 19.1 18.1 12.3 16.4 15.2
16.1 23.4 20.6 18.4 23.0 25.1 26.8 31.9
24.4 28.4 20.4 22.1 15.3 10.9 24.2 22.6

22. Power Failures The durations (in minutes) of power failures at a residence in the last 10 years

18 26 45 75 125 80 33 40 44 49
89 80 96 125 12 61 31 63 103 28

Type of lenses	Frequency, f
Contacts	40
Eyeglasses	570
Contacts and eyeglasses	180
None	210

TABLE FOR EXERCISE 23

23. Eyeglasses and Contacts The responses of a sample of 1000 adults who were asked what type of corrective lenses they wore are shown in the table at the left. *(Adapted from American Optometric Association)*

24. Living on Your Own The responses of a sample of 1177 young adults who were asked what surprised them the most as they began to live on their own *(Adapted from Charles Schwab)*

Amount of first salary: 63 Trying to find a job: 125
Number of decisions: 163 Money needed: 326
Paying bills: 150 Trying to save: 275
How hard it is breaking away from parents: 75

25. Top Speeds The top speeds (in miles per hour) for a sample of seven sports cars

187.3 181.8 180.0 169.3 162.2 158.1 155.7

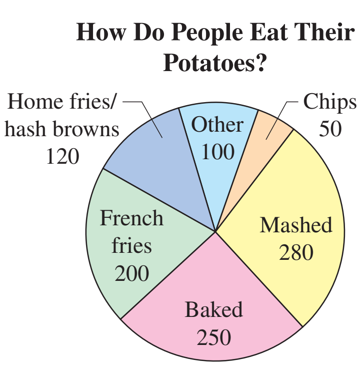


FIGURE FOR EXERCISE 26

26. Potatoes The pie chart at the left shows the responses of a sample of 1000 adults who were asked their favorite way to eat potatoes. *(Adapted from Idaho Potato Commission)*

27. Typing Speeds The typing speeds (in words per minute) for several stenographers

125 140 170 155 132 175 225 210 125 230

28. Eating Disorders The number of weeks it took to reach a target weight for a sample of five patients with eating disorders treated by psychodynamic psychotherapy *(Source: The Journal of Consulting and Clinical Psychology)*

15.0 31.5 10.0 25.5 1.0

29. Eating Disorders The number of weeks it took to reach a target weight for a sample of 14 patients with eating disorders treated by psychodynamic psychotherapy and cognitive behavior techniques *(Source: The Journal of Consulting and Clinical Psychology)*

2.5	20.0	11.0	10.5	17.5	16.5	13.0
15.5	26.5	2.5	27.0	28.5	1.5	5.0

30. Aircraft The number of aircraft that 15 airlines have in their fleets *(Source: Airline Transport Association)*

136	110	38	625	350	755	52	32
142	9	537	28	409	354	28	

31. Weights (in pounds) of Carry-On Luggage on a Plane

Key: 3|2 = 32

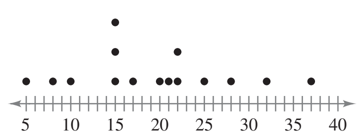
0	6 7
1	2 5 8 9
2	0 4 4 4 5 8 9
3	2 2 3 5 5 5 6 8 9
4	0 1 2 7 8
5	1

32. Grade Point Averages of Students in a Class

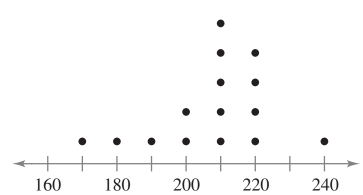
Key: 0|8 = 0.8

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

33. Time (in minutes) It Takes Employees to Drive to Work

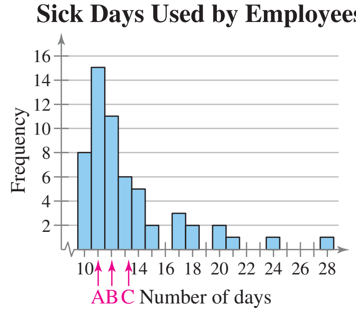


34. Prices (in dollars per night) of Hotel Rooms in a City

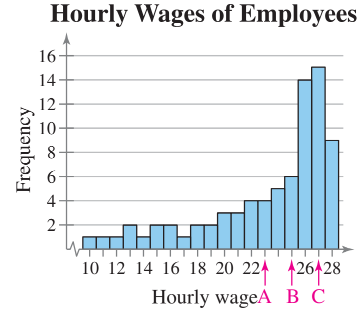


Graphical Analysis *In Exercises 35 and 36, the letters A, B, and C are marked on the horizontal axis. Describe the shape of the data. Then determine which is the mean, which is the median, and which is the mode. Justify your answers.*

35. Sick Days Used by Employees

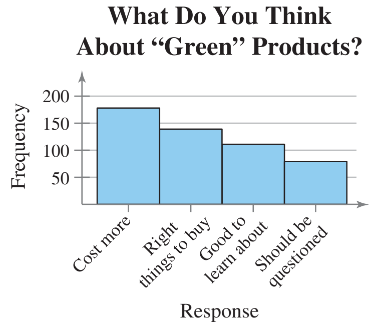


36. Hourly Wages of Employees



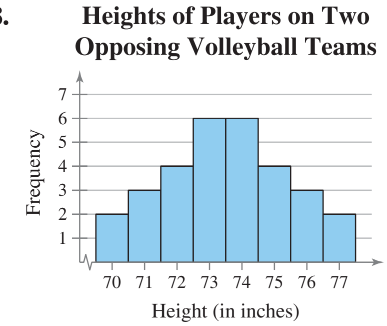
In Exercises 37–40, without performing any calculations, determine which measure of central tendency best represents the graphed data. Explain your reasoning.

37.

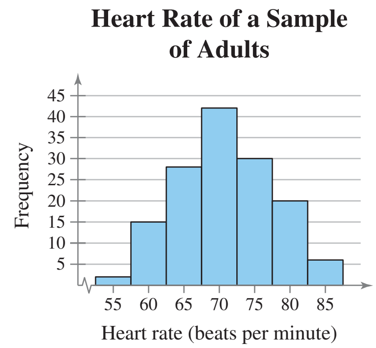


(Adapted from Green Home Furnishings Consumer Study)

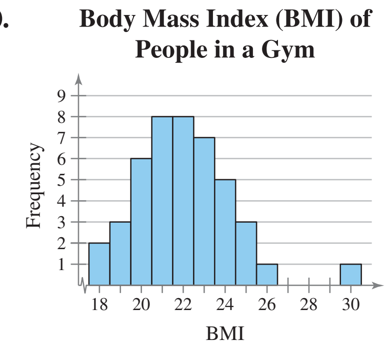
38.



39.



40.



Finding the Weighted Mean *In Exercises 41–46, find the weighted mean of the data.*

41. Final Grade The scores and their percents of the final grade for a statistics student are given. What is the student's mean score?

	Score	Percent of final grade
Homework	85	5%
Quizzes	80	35%
Project	100	20%
Speech	90	15%
Final exam	93	25%

42. Salaries The average starting salaries (by degree attained) for 25 employees at a company are given. What is the mean starting salary for these employees?

8 with MBAs: \$92,500   17 with BAs in business: \$68,000

43. Account Balance For the month of April, a checking account has a balance of \$523 for 24 days, \$2415 for 2 days, and \$250 for 4 days. What is the account's mean daily balance for April?

44. Account Balance For the month of May, a checking account has a balance of \$759 for 15 days, \$1985 for 5 days, \$1410 for 5 days, and \$348 for 6 days. What is the account's mean daily balance for May?

45. Grades A student receives the following grades, with an A worth 4 points, a B worth 3 points, a C worth 2 points, and a D worth 1 point. What is the student's mean grade point score?

B in 2 three-credit classes   D in 1 two-credit class
A in 1 four-credit class   C in 1 three-credit class

46. Scores The mean scores for students in a statistics course (by major) are given. What is the mean score for the class?

9 engineering majors: 85
5 math majors: 90
13 business majors: 81

47. Final Grade In Exercise 41, an error was made in grading your final exam. Instead of getting 93, you scored 85. What is your new weighted mean?

48. Grades In Exercise 45, one of the student's B grades gets changed to an A. What is the student's new mean grade point score?

Finding the Mean of Grouped Data *In Exercises 49–52, approximate the mean of the grouped data.*

49. Fuel Economy The highway mileage (in miles per gallon) for 30 small cars

Mileage (miles per gallon)	Frequency
29–33	11
34–38	12
39–43	2
44–48	5

50. Fuel Economy The city mileage (in miles per gallon) for 24 family sedans

Mileage (miles per gallon)	Frequency
22–27	16
28–33	2
34–39	2
40–45	3
46–51	1

51. Ages The ages of residents of a town

Age	Frequency
0–9	55
10–19	70
20–29	35
30–39	56
40–49	74
50–59	42
60–69	38
70–79	17
80–89	10

52. Phone Calls The lengths of calls (in minutes) made by a salesperson in one week

Length of call	Number of calls
1–5	12
6–10	26
11–15	20
16–20	7
21–25	11
26–30	7
31–35	4
36–40	4
41–45	1

Identifying the Shape of a Distribution *In Exercises 53–56, construct a frequency distribution and a frequency histogram of the data using the indicated number of classes. Describe the shape of the histogram as symmetric, uniform, negatively skewed, positively skewed, or none of these.*

53. Hospital Beds
Number of classes: 5
Data set: The number of beds in a sample of 24 hospitals

149	167	162	127	130	180	160	167
221	145	137	194	207	150	254	262
244	297	137	204	166	174	180	151

54. Hospitalization
Number of classes: 6
Data set: The number of days 20 patients remained hospitalized

6	9	7	14	4	5	6	8	4	11
10	6	8	6	5	7	6	6	3	11

55. Heights of Males
Number of classes: 5
Data set: The heights (to the nearest inch) of 30 males

67	76	69	68	72	68	65	63	75	69
66	72	67	66	69	73	64	62	71	73
68	72	71	65	69	66	74	72	68	69

56. Six-Sided Die
Number of classes: 6
Data set: The results of rolling a six-sided die 30 times

1	4	6	1	5	3	2	5	4	6	1	2	4	3	5
6	3	2	1	1	5	6	2	4	4	3	1	6	2	4

57. Coffee Contents During a quality assurance check, the actual coffee contents (in ounces) of six jars of instant coffee were recorded as 6.03, 5.59, 6.40, 6.00, 5.99, and 6.02.

(a) Find the mean and the median of the coffee content.

(b) The third value was incorrectly measured and is actually 6.04. Find the mean and median of the coffee content again.

(c) Which measure of central tendency, the mean or the median, was affected more by the data entry error?

58. U.S. Exports The table at the left shows the U.S. exports (in billions of dollars) to 19 countries for a recent year. *(Source: U.S. Department of Commerce)*

(a) Find the mean and median.

(b) Find the mean and median without the U.S. exports to Canada. Which measure of central tendency, the mean or the median, was affected more by the elimination of the Canadian exports?

(c) The U.S. exports to India were $17.7 billion. Find the mean and median with the Indian exports added to the original data set. Which measure of central tendency was affected more by adding the Indian exports?

U.S. Exports (in billions of dollars)	
Canada: 261.1	Japan: 65.1
Mexico: 151.2	South Korea: 34.7
Germany: 54.5	Singapore: 27.9
Taiwan: 24.9	France: 28.8
Netherlands: 39.7	Brazil: 32.3
China: 69.7	Belgium: 28.9
Australia: 22.2	Italy: 15.5
Malaysia: 12.9	Thailand: 9.1
Switzerland: 22.0	
Saudi Arabia: 12.5	
United Kingdom: 53.6	

TABLE FOR EXERCISE 58

SC *In Exercises 59 and 60, use StatCrunch to find the sample size, mean, median, minimum data value, and maximum data value of the data.*

59. The data represent the amounts (in dollars) made by several families during a community yard sale.

95 120 125.50 105.25 82 102.75 130 151.50 145.25 79 97

60. The data represent the prices (in dollars) of the stocks in the Dow Jones Industrial Average during a recent session. *(Source: CNN Money)*

83.62	15.90	42.61	26.35	16.89	61.46	62.07	79.53	24.99	34.05
69.62	16.77	52.69	21.46	132.39	65.10	44.56	29.08	62.54	39.92
31.07	19.46	57.19	28.30	61.49	49.28	72.77	31.38	54.33	31.06

EXTENDING CONCEPTS

61. Golf The distances (in yards) for nine holes of a golf course are listed.

336 393 408 522 147 504 177 375 360

(a) Find the mean and median of the data.

(b) Convert the distances to feet. Then rework part (a).

(c) Compare the measures you found in part (b) with those found in part (a). What do you notice?

(d) Use your results from part (c) to explain how to find quickly the mean and median of the given data set if the distances are measured in inches.

62. Data Analysis A consumer testing service obtained the following mileages (in miles per gallon) in five test runs performed with three types of compact cars.

	Run 1	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Car A:	28	32	28	30	34
Car B:	31	29	31	29	31
Car C:	29	32	28	32	30

(a) The manufacturer of Car A wants to advertise that its car performed best in this test. Which measure of central tendency—mean, median, or mode—should be used for its claim? Explain your reasoning.

(b) The manufacturer of Car B wants to advertise that its car performed best in this test. Which measure of central tendency—mean, median, or mode—should be used for its claim? Explain your reasoning.

(c) The manufacturer of Car C wants to advertise that its car performed best in this test. Which measure of central tendency—mean, median, or mode—should be used for its claim? Explain your reasoning.

63. Midrange Another measure of central tendency that is rarely used but is easy to calculate is the **midrange.** It can be found by the formula

$$\frac{(\text{Maximum data entry}) + (\text{Minimum data entry})}{2}.$$

Which of the manufacturers in Exercise 62 would prefer to use the midrange statistic in their ads? Explain your reasoning.

64. Data Analysis Students in an experimental psychology class did research on depression as a sign of stress. A test was administered to a sample of 30 students. The scores are given.

44 51 11 90 76 36 64 37 43 72 53 62 36 74 51
72 37 28 38 61 47 63 36 41 22 37 51 46 85 13

(a) Find the mean and median of the data.

(b) Draw a stem-and-leaf plot for the data using one row per stem. Locate the mean and median on the display.

(c) Describe the shape of the distribution.

65. Trimmed Mean To find the 10% **trimmed mean** of a data set, order the data, delete the lowest 10% of the entries and the highest 10% of the entries, and find the mean of the remaining entries.

(a) Find the 10% trimmed mean for the data in Exercise 64.

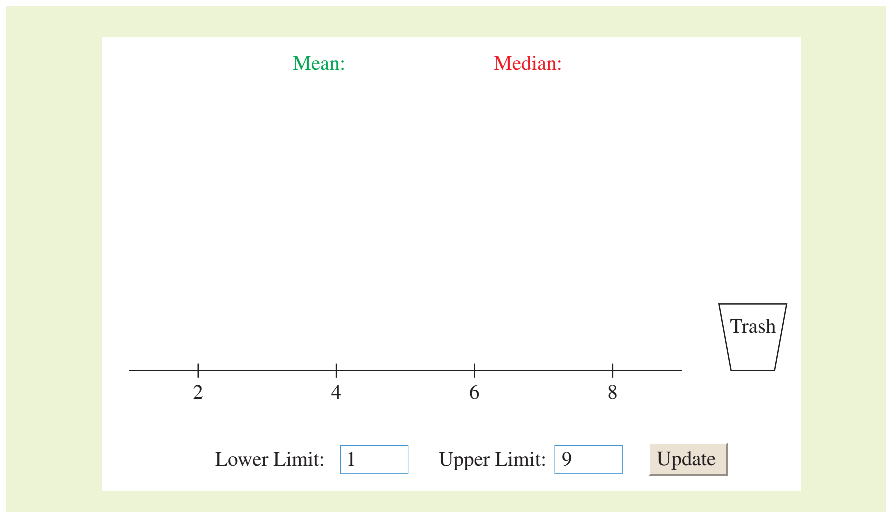
(b) Compare the four measures of central tendency, including the midrange.

(c) What is the benefit of using a trimmed mean versus using a mean found using all data entries? Explain your reasoning.

ACTIVITY 2.3 Mean Versus Median

APPLET

The *mean versus median* applet is designed to allow you to investigate interactively the mean and the median as measures of the center of a data set. Points can be added to the plot by clicking the mouse above the horizontal axis. The mean of the points is shown as a green arrow and the median is shown as a red arrow. If the two values are the same, then a single yellow arrow is displayed. Numeric values for the mean and median are shown above the plot. Points on the plot can be removed by clicking on the point and then dragging the point into the trash can. All of the points on the plot can be removed by simply clicking inside the trash can. The range of values for the horizontal axis can be specified by inputting lower and upper limits and then clicking UPDATE.



Explore

Step 1 Specify a lower limit.
Step 2 Specify an upper limit.
Step 3 Add 15 points to the plot.
Step 4 Remove all of the points from the plot.

Draw Conclusions

APPLET

1. Specify the lower limit to be 1 and the upper limit to be 50. Add at least 10 points that range from 20 to 40 so that the mean and the median are the same. What is the shape of the distribution? What happens at first to the mean and median when you add a few points that are less than 10? What happens over time as you continue to add points that are less than 10?

2. Specify the lower limit to be 0 and the upper limit to be 0.75. Place 10 points on the plot. Then change the upper limit to 25. Add 10 more points that are greater than 20 to the plot. Can the mean be any one of the points that were plotted? Can the median be any one of the points that were plotted? Explain.

2.4 Measures of Variation

WHAT YOU SHOULD LEARN

- How to find the range of a data set
- How to find the variance and standard deviation of a population and of a sample
- How to use the Empirical Rule and Chebychev's Theorem to interpret standard deviation
- How to approximate the sample standard deviation for grouped data

Range ▸ Deviation, Variance, and Standard Deviation ▸ Interpreting Standard Deviation ▸ Standard Deviation for Grouped Data

▸ RANGE

In this section, you will learn different ways to measure the variation of a data set. The simplest measure is the *range* of the set.

DEFINITION

The **range** of a data set is the difference between the maximum and minimum data entries in the set. To find the range, the data must be quantitative.

$$\text{Range} = (\text{Maximum data entry}) - (\text{Minimum data entry})$$

EXAMPLE 1

SC Report 12

▸ Finding the Range of a Data Set

Two corporations each hired 10 graduates. The starting salaries for each graduate are shown. Find the range of the starting salaries for Corporation A.

Starting Salaries for Corporation A (1000s of dollars)

Salary	41	38	39	45	47	41	44	41	37	42

Starting Salaries for Corporation B (1000s of dollars)

Salary	40	23	41	50	49	32	41	29	52	58

▸ Solution

Ordering the data helps to find the least and greatest salaries.

37 38 39 41 41 41 42 44 45 47

Minimum (37) Maximum (47)

$$\begin{aligned}\text{Range} &= (\text{Maximum salary}) - (\text{Minimum salary})\\ &= 47 - 37\\ &= 10\end{aligned}$$

So, the range of the starting salaries for Corporation A is 10, or $10,000.

▸ Try It Yourself 1

Find the range of the starting salaries for Corporation B.

a. Identify the *minimum* and *maximum* salaries.
b. Find the *range*.
c. *Compare* your answer with that for Example 1.

Answer: Page A33

INSIGHT

Both data sets in Example 1 have a mean of 41.5, or $41,500, a median of 41, or $41,000, and a mode of 41, or $41,000. And yet the two sets differ significantly.

The difference is that the entries in the second set have greater variation. Your goal in this section is to learn how to measure the variation of a data set.

DEVIATION, VARIANCE, AND STANDARD DEVIATION

As a measure of variation, the range has the advantage of being easy to compute. Its disadvantage, however, is that it uses only two entries from the data set. Two measures of variation that use all the entries in a data set are the *variance* and the *standard deviation*. However, before you learn about these measures of variation, you need to know what is meant by the *deviation* of an entry in a data set.

DEFINITION

The **deviation** of an entry x in a population data set is the difference between the entry and the mean μ of the data set.

$$\text{Deviation of } x = x - \mu$$

EXAMPLE 2

Finding the Deviations of a Data Set

Find the deviation of each starting salary for Corporation A given in Example 1.

Solution

The mean starting salary is $\mu = 415/10 = 41.5$, or \$41,500. To find out how much each salary deviates from the mean, subtract 41.5 from the salary. For instance, the deviation of 41, or \$41,000 is

$$41 - 41.5 = -0.5, \text{ or } -\$500. \qquad \text{Deviation of } x = x - \mu$$

x μ

The table at the left lists the deviations of each of the 10 starting salaries.

Deviations of Starting Salaries for Corporation A

Salary (1000s of dollars) x	Deviation (1000s of dollars) $x - \mu$
41	−0.5
38	−3.5
39	−2.5
45	3.5
47	5.5
41	−0.5
44	2.5
41	−0.5
37	−4.5
42	0.5
$\Sigma x = 415$	$\Sigma(x - \mu) = 0$

Try It Yourself 2

Find the deviation of each starting salary for Corporation B given in Example 1.

a. Find the *mean* of the data set.
b. *Subtract* the mean from each salary.

Answer: Page A33

In Example 2, notice that the sum of the deviations is zero. Because this is true for any data set, it doesn't make sense to find the average of the deviations. To overcome this problem, you can square each deviation. When you add the squares of the deviations, you compute a quantity called the **sum of squares,** denoted SS_x. In a population data set, the mean of the squares of the deviations is called the **population variance.**

DEFINITION

The **population variance** of a population data set of N entries is

$$\text{Population variance} = \sigma^2 = \frac{\Sigma(x - \mu)^2}{N}.$$

The symbol σ is the lowercase Greek letter sigma.

DEFINITION

The **population standard deviation** of a population data set of N entries is the square root of the population variance.

$$\text{Population standard deviation} = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$$

GUIDELINES

Finding the Population Variance and Standard Deviation

IN WORDS	IN SYMBOLS
1. Find the mean of the population data set.	$\mu = \frac{\Sigma x}{N}$
2. Find the deviation of each entry.	$x - \mu$
3. Square each deviation.	$(x-\mu)^2$
4. Add to get the **sum of squares.**	$SS_x = \Sigma(x-\mu)^2$
5. Divide by N to get the **population variance.**	$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$
6. Find the square root of the variance to get the **population standard deviation.**	$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

Sum of Squares of Starting Salaries for Corporation A

Salary x	Deviation $x - \mu$	Squares $(x-\mu)^2$
41	−0.5	0.25
38	−3.5	12.25
39	−2.5	6.25
45	3.5	12.25
47	5.5	30.25
41	−0.5	0.25
44	2.5	6.25
41	−0.5	0.25
37	−4.5	20.25
42	0.5	0.25
	$\Sigma = 0$	$SS_x = 88.5$

EXAMPLE 3

Finding the Population Standard Deviation

Find the population standard deviation of the starting salaries for Corporation A given in Example 1.

Solution

The table at the left summarizes the steps used to find SS_x.

$$SS_x = 88.5, \qquad N = 10, \qquad \sigma^2 = \frac{88.5}{10} \approx 8.9, \qquad \sigma = \sqrt{\frac{88.5}{10}} \approx 3.0$$

So, the population variance is about 8.9, and the population standard deviation is about 3.0, or $3000.

Try It Yourself 3

Find the population variance and standard deviation of the starting salaries for Corporation B given in Example 1.

a. Find the *mean* and each *deviation*, as you did in Try It Yourself 2.
b. *Square* each deviation and *add* to get the sum of squares.
c. *Divide* by N to get the population variance.
d. Find the *square root* of the population variance to get the population standard deviation.
e. *Interpret* the results by giving the population standard deviation in dollars.

Answer: Page A33

STUDY TIP

Notice that the variance and standard deviation in Example 3 have one more decimal place than the original set of data values has. This is the same *round-off rule* that was used to calculate the mean.

STUDY TIP

Note that when you find the *population* variance, you divide by N, the number of entries, but, for technical reasons, when you find the *sample* variance, you divide by $n - 1$, one less than the number of entries.

Symbols in Variance and Standard Deviation Formulas

	Population	Sample
Variance	σ^2	s^2
Standard deviation	σ	s
Mean	μ	$\bar{x}$
Number of entries	N	n
Deviation	$x - \mu$	$x - \bar{x}$
Sum of squares	$\Sigma(x - \mu)^2$	$\Sigma(x - \bar{x})^2$

See MINITAB and TI-83/84 Plus steps on pages 122 and 123.

DEFINITION

The **sample variance** and **sample standard deviation** of a sample data set of n entries are listed below.

$$\text{Sample variance} = s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$$

$$\text{Sample standard deviation} = s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

GUIDELINES

Finding the Sample Variance and Standard Deviation

IN WORDS	IN SYMBOLS
1. Find the mean of the sample data set.	$\bar{x} = \frac{\Sigma x}{n}$
2. Find the deviation of each entry.	$x - \bar{x}$
3. Square each deviation.	$(x - \bar{x})^2$
4. Add to get the **sum of squares.**	$SS_x = \Sigma(x - \bar{x})^2$
5. Divide by $n - 1$ to get the **sample variance.**	$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$
6. Find the square root of the variance to get the **sample standard deviation.**	$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$

EXAMPLE 4

SC Report 13

Finding the Sample Standard Deviation

The starting salaries given in Example 1 are for the Chicago branches of Corporations A and B. Each corporation has several other branches, and you plan to use the starting salaries of the Chicago branches to estimate the starting salaries for the larger populations. Find the *sample* standard deviation of the starting salaries for the Chicago branch of Corporation A.

Solution

$$SS_x = 88.5, \quad n = 10, \quad s^2 = \frac{88.5}{9} \approx 9.8, \quad s = \sqrt{\frac{88.5}{9}} \approx 3.1$$

So, the sample variance is about 9.8, and the sample standard deviation is about 3.1, or \$3100.

Try It Yourself 4

Find the sample standard deviation of the starting salaries for the Chicago branch of Corporation B.

a. Find the *sum of squares*, as you did in Try It Yourself 3.
b. *Divide* by $n - 1$ to get the sample variance.
c. Find the *square root* of the sample variance to get the sample standard deviation.
d. *Interpret* the results by giving the sample standard deviation in dollars.

Answer: Page A33

EXAMPLE 5

▸ Using Technology to Find the Standard Deviation

Sample office rental rates (in dollars per square foot per year) for Miami's central business district are shown in the table. Use a calculator or a computer to find the mean rental rate and the sample standard deviation. *(Adapted from Cushman & Wakefield Inc.)*

Office Rental Rates		
35.00	33.50	37.00
23.75	26.50	31.25
36.50	40.00	32.00
39.25	37.50	34.75
37.75	37.25	36.75
27.00	35.75	26.00
37.00	29.00	40.50
24.50	33.00	38.00

▸ Solution

MINITAB, Excel, and the TI-83/84 Plus each have features that automatically calculate the means and the standard deviations of data sets. Try using this technology to find the mean and the standard deviation of the office rental rates. From the displays, you can see that $\bar{x} \approx 33.73$ and $s \approx 5.09$.

MINITAB

Descriptive Statistics: Rental Rates

Variable	N	Mean	SE Mean	StDev	Minimum
Rental Rates	24	33.73	1.04	5.09	23.75

Variable	Q1	Median	Q3	Maximum
Rental Rates	29.56	35.38	37.44	40.50

EXCEL

	A	B
1	Mean	33.72917
2	Standard Error	1.038864
3	Median	35.375
4	Mode	37
5	Standard Deviation	5.089373
6	Sample Variance	25.90172
7	Kurtosis	-0.74282
8	Skewness	-0.70345
9	Range	16.75
10	Minimum	23.75
11	Maximum	40.5
12	Sum	809.5
13	Count	24

TI-83/84 PLUS

1-Var Stats
$\bar{x}$=33.72916667
Σx=809.5
Σx^2=27899.5
Sx=5.089373342
σx=4.982216639
n=24

Sample Mean
Sample Standard Deviation

STUDY TIP

Here are instructions for calculating the sample mean and sample standard deviation on a TI-83/84 Plus for Example 5.

STAT

Choose the EDIT menu.

1: Edit

Enter the sample office rental rates into L1.

STAT

Choose the CALC menu.

1: 1–Var Stats

ENTER

2nd L1 ENTER

▸ Try It Yourself 5

Sample office rental rates (in dollars per square foot per year) for Seattle's central business district are listed. Use a calculator or a computer to find the mean rental rate and the sample standard deviation. *(Adapted from Cushman & Wakefield Inc.)*

40.00	43.00	46.00	40.50	35.75	39.75	32.75
36.75	35.75	38.75	38.75	36.75	38.75	39.00
29.00	35.00	42.75	32.75	40.75	35.25	

a. *Enter* the data.
b. *Calculate* the sample mean and the sample standard deviation.

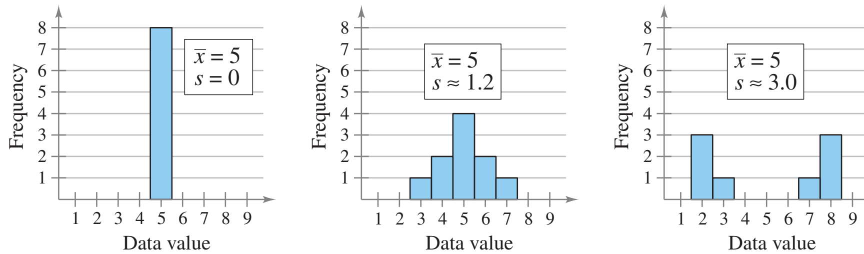
Answer: Page A33

INSIGHT

When all data values are equal, the standard deviation is 0. Otherwise, the standard deviation must be positive.

▶ INTERPRETING STANDARD DEVIATION

When interpreting the standard deviation, remember that it is a measure of the typical amount an entry deviates from the mean. The more the entries are spread out, the greater the standard deviation.

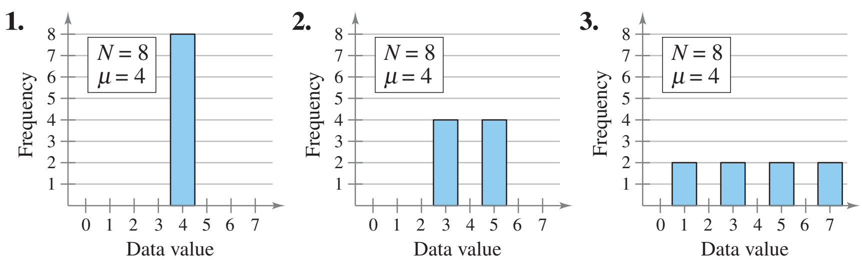


EXAMPLE 6

To explore this topic further, see Activity 2.4 on page 98.

▶ Estimating Standard Deviation

Without calculating, estimate the population standard deviation of each data set.



▶ Solution

1. Each of the eight entries is 4. So, each deviation is 0, which implies that

   $$\sigma = 0.$$

2. Each of the eight entries has a deviation of ± 1. So, the population standard deviation should be 1. By calculating, you can see that

   $$\sigma = 1.$$

3. Each of the eight entries has a deviation of ± 1 or ± 3. So, the population standard deviation should be about 2. By calculating, you can see that

   $$\sigma \approx 2.24.$$

▶ Try It Yourself 6

Write a data set that has 10 entries, a mean of 10, and a population standard deviation that is approximately 3. (There are many correct answers.)

a. *Write* a data set that has five entries that are three units less than 10 and five entries that are three units more than 10.

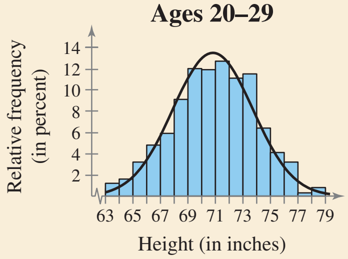
b. *Calculate* the population standard deviation to check that σ is approximately 3.

Answer: Page A33

PICTURING THE WORLD

A survey was conducted by the National Center for Health Statistics to find the mean height of males in the United States. The histogram shows the distribution of heights for the 808 men examined in the 20–29 age group. In this group, the mean was 69.9 inches and the standard deviation was 3.0 inches. (Adapted from National Center for Health Statistics)

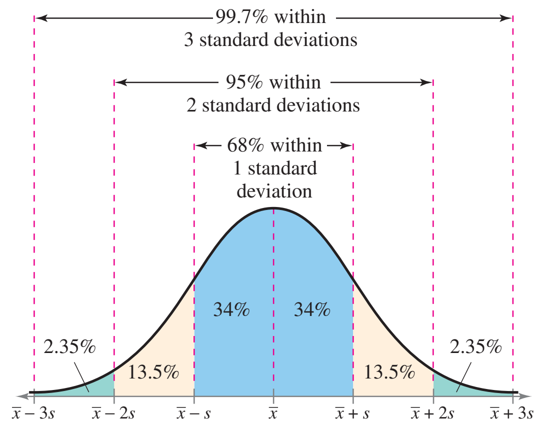
Heights of Men in the U.S. Ages 20–29



Roughly which two heights contain the middle 95% of the data?

Many real-life data sets have distributions that are approximately symmetric and bell-shaped. Later in the text, you will study this type of distribution in detail. For now, however, the following *Empirical Rule* can help you see how valuable the standard deviation can be as a measure of variation.

Bell-Shaped Distribution



EMPIRICAL RULE (OR 68–95–99.7 RULE)

For data with a (symmetric) bell-shaped distribution, the standard deviation has the following characteristics.

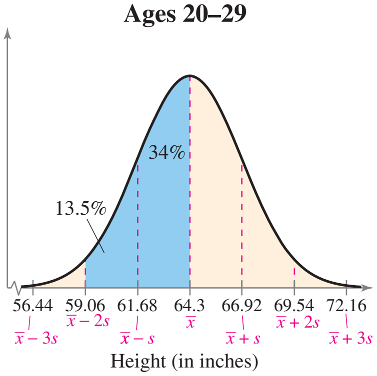
1. About 68% of the data lie within one standard deviation of the mean.
2. About 95% of the data lie within two standard deviations of the mean.
3. About 99.7% of the data lie within three standard deviations of the mean.

EXAMPLE 7

Using the Empirical Rule

In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20–29) was 64.3 inches, with a sample standard deviation of 2.62 inches. Estimate the percent of women whose heights are between 59.06 inches and 64.3 inches. *(Adapted from National Center for Health Statistics)*

Heights of Women in the U.S. Ages 20–29



Solution

The distribution of women's heights is shown. Because the distribution is bell-shaped, you can use the Empirical Rule. The mean height is 64.3, so when you subtract two standard deviations from the mean height, you get

$$\bar{x} - 2s = 64.3 - 2(2.62) = 59.06.$$

Because 59.06 is two standard deviations below the mean height, the percent of the heights between 59.06 and 64.3 inches is $13.5\% + 34\% = 47.5\%$.

Interpretation So, 47.5% of women are between 59.06 and 64.3 inches tall.

Try It Yourself 7

Estimate the percent of women's heights that are between 64.3 and 66.92 inches tall.

a. How many *standard deviations* is 66.92 to the right of 64.3?

b. Use the *Empirical Rule* to estimate the percent of the data between $\bar{x}$ and $\bar{x} + s$.

c. *Interpret* the result in the context of the data.

Answer: Page A33

INSIGHT

Data values that lie more than two standard deviations from the mean are considered unusual. Data values that lie more than three standard deviations from the mean are very unusual.

The Empirical Rule applies only to (symmetric) bell-shaped distributions. What if the distribution is not bell-shaped, or what if the shape of the distribution is not known? The following theorem gives an inequality statement that applies to *all* distributions. It is named after the Russian statistician Pafnuti Chebychev (1821–1894).

CHEBYCHEV'S THEOREM

The portion of any data set lying within k standard deviations $(k > 1)$ of the mean is at least

$$1 - \frac{1}{k^2}.$$

- $k = 2$: In any data set, at least $1 - \frac{1}{2^2} = \frac{3}{4}$, or 75%, of the data lie within 2 standard deviations of the mean.
- $k = 3$: In any data set, at least $1 - \frac{1}{3^2} = \frac{8}{9}$, or 88.9%, of the data lie within 3 standard deviations of the mean.

INSIGHT

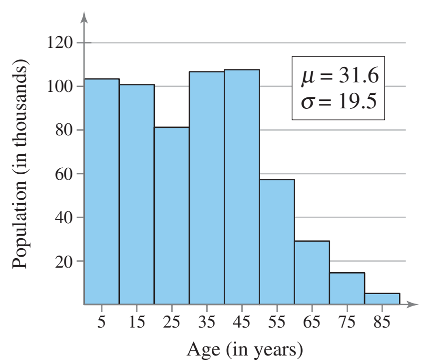
In Example 8, Chebychev's Theorem gives you an inequality statement that says that at least 75% of the population of Florida is under the age of 88.8. This is a true statement, but it is not nearly as strong a statement as could be made from reading the histogram.

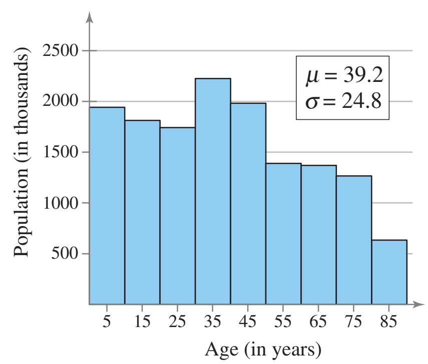
In general, Chebychev's Theorem gives the minimum percent of data values that fall within the given number of standard deviations of the mean. Depending on the distribution, there is probably a higher percent of data falling in the given range.

EXAMPLE 8

Using Chebychev's Theorem

The age distributions for Alaska and Florida are shown in the histograms. Decide which is which. Apply Chebychev's Theorem to the data for Florida using $k = 2$. What can you conclude?





Solution

The histogram on the right shows Florida's age distribution. You can tell because the population is greater and older. Moving two standard deviations to the left of the mean puts you below 0, because $\mu - 2\sigma = 39.2 - 2(24.8) = -10.4$. Moving two standard deviations to the right of the mean puts you at $\mu + 2\sigma = 39.2 + 2(24.8) = 88.8$. By Chebychev's Theorem, you can say that at least 75% of the population of Florida is between 0 and 88.8 years old.

Try It Yourself 8

Apply Chebychev's Theorem to the data for Alaska using $k = 2$. What can you conclude?

a. *Subtract* two standard deviations from the mean.
b. *Add* two standard deviations to the mean.
c. *Apply* Chebychev's Theorem for $k = 2$ and *interpret* the results.

Answer: Page A33

STUDY TIP

Remember that formulas for grouped data require you to multiply by the frequencies.

STANDARD DEVIATION FOR GROUPED DATA

In Section 2.1, you learned that large data sets are usually best represented by frequency distributions. The formula for the sample standard deviation for a frequency distribution is

$$\text{Sample standard deviation} = s = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{n-1}}$$

where $n = \Sigma f$ is the number of entries in the data set.

EXAMPLE 9

Finding the Standard Deviation for Grouped Data

You collect a random sample of the number of children per household in a region. The results are shown at the left. Find the sample mean and the sample standard deviation of the data set.

Number of Children in 50 Households				
1	3	1	1	1
1	2	2	1	0
1	1	0	0	0
1	5	0	3	6
3	0	3	1	1
1	1	6	0	1
3	6	6	1	2
2	3	0	1	1
4	1	1	2	2
0	3	0	2	4

Solution

These data could be treated as 50 individual entries, and you could use the formulas for mean and standard deviation. Because there are so many repeated numbers, however, it is easier to use a frequency distribution.

x	f	xf
0	10	0
1	19	19
2	7	14
3	7	21
4	2	8
5	1	5
6	4	24
	$\Sigma = 50$	$\Sigma = 91$

$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
−1.8	3.24	32.40
−0.8	0.64	12.16
0.2	0.04	0.28
1.2	1.44	10.08
2.2	4.84	9.68
3.2	10.24	10.24
4.2	17.64	70.56
		$\Sigma = 145.40$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{91}{50} \approx 1.8$$ Sample mean

Use the sum of squares to find the sample standard deviation.

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{n-1}} = \sqrt{\frac{145.4}{49}} \approx 1.7$$ Sample standard deviation

So, the sample mean is about 1.8 children, and the sample standard deviation is about 1.7 children.

Try It Yourself 9

Change three of the 6's in the data set to 4's. How does this change affect the sample mean and sample standard deviation?

a. Write the first three columns of a *frequency distribution*.
b. Find the *sample mean*.
c. Complete the *last three columns* of the frequency distribution.
d. Find the *sample standard deviation*.

Answer: Page A33

STUDY TIP

Here are instructions for calculating the sample mean and sample standard deviation on a TI-83/84 Plus for the grouped data in Example 9.

STAT

Choose the EDIT menu.

1: Edit

Enter the values of *x* into L1.
Enter the frequencies *f* into L2.

STAT

Choose the CALC menu.

1: 1–Var Stats

ENTER

2nd L1 , 2nd L2

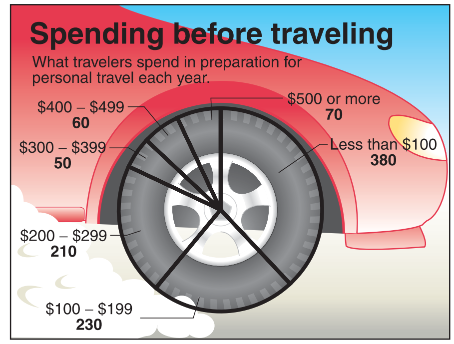
ENTER

When a frequency distribution has classes, you can estimate the sample mean and the sample standard deviation by using the midpoint of each class.

EXAMPLE 10

▸ Using Midpoints of Classes

The circle graph at the right shows the results of a survey in which 1000 adults were asked how much they spend in preparation for personal travel each year. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set. *(Adapted from Travel Industry Association of America)*



▸ Solution

Begin by using a frequency distribution to organize the data.

Class	x	f	xf
0–99	49.5	380	18,810
100–199	149.5	230	34,385
200–299	249.5	210	52,395
300–399	349.5	50	17,475
400–499	449.5	60	26,970
500+	599.5	70	41,965
		$\Sigma = 1000$	$\Sigma = 192{,}000$

$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
−142.5	20,306.25	7,716,375.0
−42.5	1806.25	415,437.5
57.5	3306.25	694,312.5
157.5	24,806.25	1,240,312.5
257.5	66,306.25	3,978,375.0
407.5	166,056.25	11,623,937.5
		$\Sigma = 25{,}668{,}750.0$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{192{,}000}{1000} = 192$$ Sample mean

Use the sum of squares to find the sample standard deviation.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2 f}{n - 1}} = \sqrt{\frac{25{,}668{,}750}{999}} \approx 160.3$$ Sample standard deviation

So, the sample mean is \$192 per year, and the sample standard deviation is about \$160.30 per year.

STUDY TIP

When a class is open, as in the last class, you must assign a single value to represent the midpoint. For this example, we selected 599.5.

▸ Try It Yourself 10

In the frequency distribution, 599.5 was chosen to represent the class of \$500 or more. How would the sample mean and standard deviation change if you used 650 to represent this class?

a. Write the first four columns of a *frequency distribution*.
b. Find the *sample mean*.
c. Complete the *last three columns* of the frequency distribution.
d. Find the *sample standard deviation*.

Answer: Page A34

2.4 EXERCISES

BUILDING BASIC SKILLS AND VOCABULARY

1. Explain how to find the range of a data set. What is an advantage of using the range as a measure of variation? What is a disadvantage?

2. Explain how to find the deviation of an entry in a data set. What is the sum of all the deviations in any data set?

3. Why is the standard deviation used more frequently than the variance? (*Hint:* Consider the units of the variance.)

4. Explain the relationship between variance and standard deviation. Can either of these measures be negative? Explain.

5. Construct a sample data set for which $n = 7$, $\bar{x} = 9$, and $s = 0$.

6. Construct a population data set for which $N = 6$, $\mu = 5$, and $\sigma = 2$.

7. Describe the difference between the calculation of population standard deviation and that of sample standard deviation.

8. Given a data set, how do you know whether to calculate σ or s?

9. Discuss the similarities and the differences between the Empirical Rule and Chebychev's Theorem.

10. What must you know about a data set before you can use the Empirical Rule?

In Exercises 11 and 12, find the range, mean, variance, and standard deviation of the population data set.

11. 9 5 9 10 11 12 7 7 8 12

12. 18 20 19 21 19 17 15
17 25 22 19 20 16 18

In Exercises 13 and 14, find the range, mean, variance, and standard deviation of the sample data set.

13. 4 15 9 12 16 8 11 19 14

14. 28 25 21 15 7 14 9
27 21 24 14 17 16

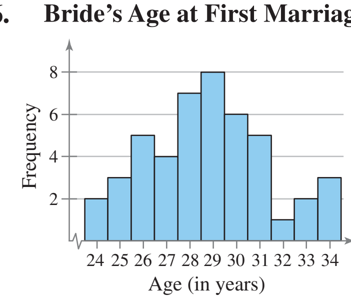
Graphical Reasoning *In Exercises 15–18, find the range of the data set represented by the display or graph.*

15.

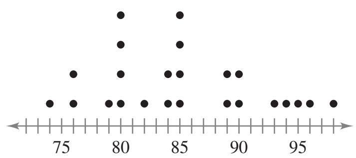
2	3 9
3	0 0 2 3 6 7
4	0 1 2 3 3 8
5	0 1 1 9
6	1 2 9 9
7	5 9
8	4 8
9	0 2 5 6

Key: $2|3 = 23$

16. **Bride's Age at First Marriage**



17.



18.

0	5 5 9	Key: 0|5 = 0.5
1	1 3 4 6 9	
2	2 5 7 9 9	
3	0 1 5 5 5 5	
4	7 7 9	
5		
6	3 4 7	

19. Archaeology The depths (in inches) at which 10 artifacts are found are given below.

20.7 24.8 30.5 26.2 36.0 34.3 30.3 29.5 27.0 38.5

(a) Find the range of the data set.

(b) Change 38.5 to 60.5 and find the range of the new data set.

20. In Exercise 19, compare your answer to part (a) with your answer to part (b). How do outliers affect the range of a data set?

USING AND INTERPRETING CONCEPTS

21. Graphical Reasoning Both data sets have a mean of 165. One has a standard deviation of 16, and the other has a standard deviation of 24. By looking at the graphs, which is which? Explain your reasoning.

(a)

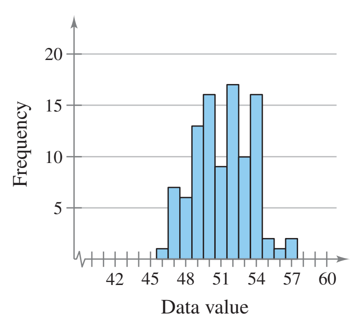
12	8 9	Key: 12|8 = 128
13	5 5 8	
14	1 2	
15	0 0 6 7	
16	4 5 9	
17	1 3 6 8	
18	0 8 9	
19	6	
20	3 5 7	

(b)

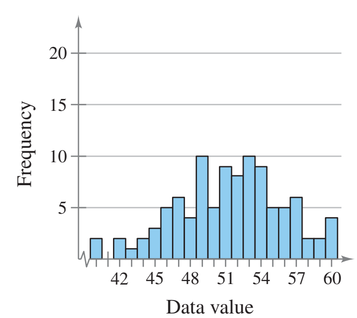
12		Key: 13|1 = 131
13	1	
14	2 3 5	
15	0 4 5 6 8	
16	1 1 2 3 3 3	
17	1 5 8 8	
18	2 3 4 5	
19	0 2	
20		

22. Graphical Reasoning Both data sets represented below have a mean of 50. One has a standard deviation of 2.4, and the other has a standard deviation of 5. By looking at the graphs, which is which? Explain your reasoning.

(a)



(b)



23. Salary Offers You are applying for jobs at two companies. Company A offers starting salaries with $\mu = \$31{,}000$ and $\sigma = \$1000$. Company B offers starting salaries with $\mu = \$31{,}000$ and $\sigma = \$5000$. From which company are you more likely to get an offer of \$33,000 or more? Explain your reasoning.

24. Golf Strokes An Internet site compares the strokes per round for two professional golfers. Which golfer is more consistent: Player A with $\mu = 71.5$ strokes and $\sigma = 2.3$ strokes, or Player B with $\mu = 70.1$ strokes and $\sigma = 1.2$ strokes? Explain your reasoning.

Comparing Two Data Sets *In Exercises 25–28, you are asked to compare two data sets and interpret the results.*

25. Annual Salaries Sample annual salaries (in thousands of dollars) for accountants in Dallas and New York City are listed.

Dallas:	41.6	50.0	49.5	38.7	39.9	45.8	44.7	47.8	40.5
New York City:	45.6	41.5	57.6	55.1	59.3	59.0	50.6	47.2	42.3

(a) Find the mean, median, range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

26. Annual Salaries Sample annual salaries (in thousands of dollars) for electrical engineers in Boston and Chicago are listed.

Boston:	70.4	84.2	58.5	64.5	71.6	79.9	88.3	80.1	69.9
Chicago:	69.4	71.5	65.4	59.9	70.9	68.5	62.9	70.1	60.9

(a) Find the mean, median, range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

27. SAT Scores Sample SAT scores for eight males and eight females are listed.

Male SAT scores:	1520	1750	2120	1380	1982	1645	1033	1714
Female SAT scores:	1785	1507	1497	1952	2210	1871	1263	1588

(a) Find the mean, median, range, variance, and standard deviation of each data set.

(b) Interpret the results in the context of the real-life setting.

28. Batting Averages Sample batting averages for baseball players from two opposing teams are listed.

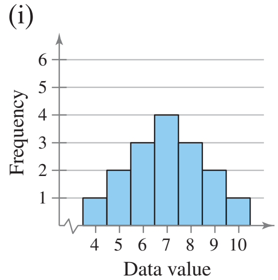
Team A:	0.295	0.310	0.325	0.272	0.256	0.297	0.320	0.384	0.235
Team B:	0.285	0.305	0.315	0.270	0.292	0.330	0.335	0.268	0.290

(a) Find the mean, median, range, variance, and standard deviation of each data set.

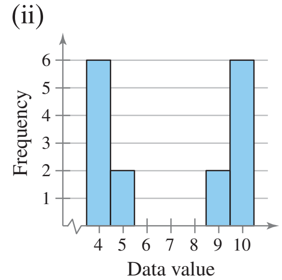
(b) Interpret the results in the context of the real-life setting.

Reasoning with Graphs *In Exercises 29–32, you are asked to compare three data sets. (a) Without calculating, determine which data set has the greatest sample standard deviation and which has the least sample standard deviation. Explain your reasoning. (b) How are the data sets the same? How do they differ?*

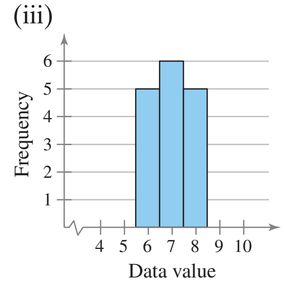
29. (i)



(ii)



(iii)



30. (i)

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

Key: 1|5 = 15

(ii)

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

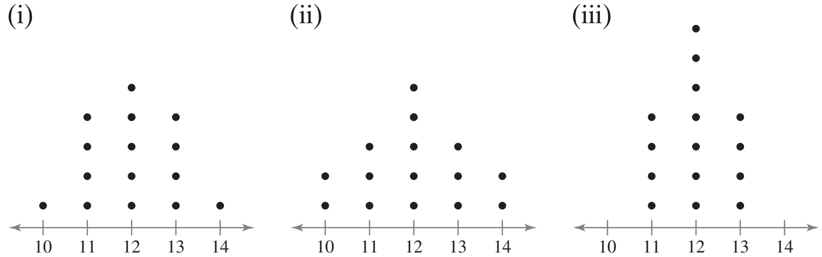
Key: 1|5 = 15

(iii)

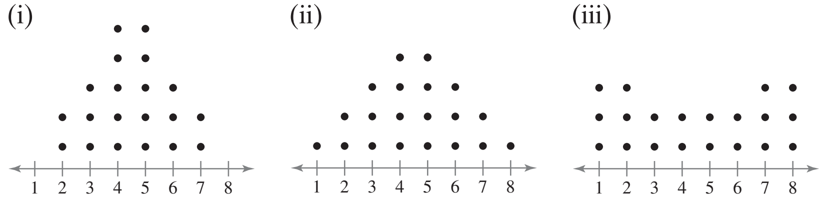
0	
1	5
2	3 3 3 3 7 7 7 7
3	5
4	

Key: 1|5 = 15

31. (i) (ii) (iii)



32. (i) (ii) (iii)



Using the Empirical Rule *In Exercises 33–38, you are asked to use the Empirical Rule.*

33. The mean value of land and buildings per acre from a sample of farms is \$1500, with a standard deviation of \$200. Estimate the percent of farms whose land and building values per acre are between \$1300 and \$1700. (Assume the data set has a bell-shaped distribution.)

34. The mean value of land and buildings per acre from a sample of farms is \$2400, with a standard deviation of \$450. Between what two values do about 95% of the data lie? (Assume the data set has a bell-shaped distribution.)

35. Using the sample statistics from Exercise 33, do the following. (Assume the number of farms in the sample is 75.)

(a) Estimate the number of farms whose land and building values per acre are between \$1300 and \$1700.

(b) If 25 additional farms were sampled, about how many of these farms would you expect to have land and building values between \$1300 per acre and \$1700 per acre?

36. Using the sample statistics from Exercise 34, do the following. (Assume the number of farms in the sample is 40.)

(a) Estimate the number of farms whose land and building values per acre are between \$1500 and \$3300.

(b) If 20 additional farms were sampled, about how many of these farms would you expect to have land and building values between \$1500 per acre and \$3300 per acre?

37. The land and building values per acre for eight more farms are listed. Using the sample statistics from Exercise 33, determine which of the data values are unusual. Are any of the data values very unusual? Explain.

\$1150, \$1775, \$2180, \$1000, \$1475, \$2000, \$1850, \$950

38. The land and building values per acre for eight more farms are listed. Using the sample statistics from Exercise 34, determine which of the data values are unusual. Are any of the data values very unusual? Explain.

$3325, $1045, $2450, $3200, $3800, $1490, $1675, $2950

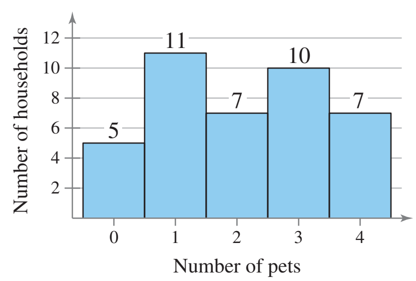
39. Chebychev's Theorem Old Faithful is a famous geyser at Yellowstone National Park. From a sample with $n = 32$, the mean duration of Old Faithful's eruptions is 3.32 minutes and the standard deviation is 1.09 minutes. Using Chebychev's Theorem, determine at least how many of the eruptions lasted between 1.14 minutes and 5.5 minutes. *(Source: Yellowstone National Park)*

40. Chebychev's Theorem The mean time in a women's 400-meter dash is 57.07 seconds, with a standard deviation of 1.05 seconds. Apply Chebychev's Theorem to the data using $k = 2$. Interpret the results.

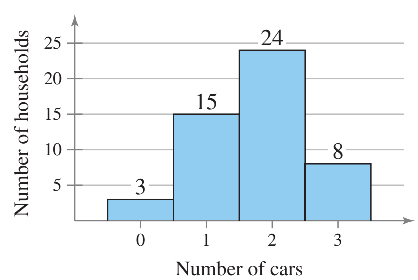
Calculating Using Grouped Data

In Exercises 41–48, use the grouped data formulas to find the indicated mean and standard deviation.

41. Pets per Household The results of a random sample of the number of pets per household in a region are shown in the histogram. Estimate the sample mean and the sample standard deviation of the data set.



42. Cars per Household The results of a random sample of the number of cars per household in a region are shown in the histogram. Estimate the sample mean and the sample standard deviation of the data set.



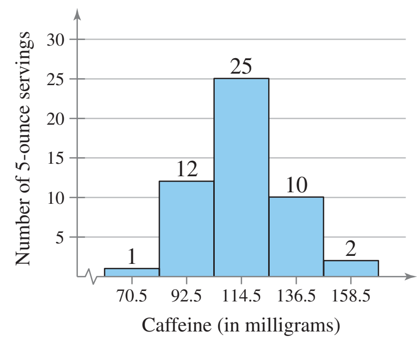
43. Football Wins The number of regular season wins for each National Football League team in 2009 are listed. Make a frequency distribution (using five classes) for the data set. Then approximate the population mean and the population standard deviation of the data set. *(Source: National Football League)*

10	9	7	6	10	9	9	5	14	9	8	7
13	8	5	4	11	11	8	4	12	11	7	2
13	9	8	3	10	8	5	1				

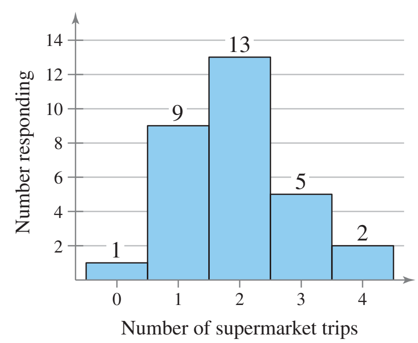
44. Water Consumption The number of gallons of water consumed per day by a small village are listed. Make a frequency distribution (using five classes) for the data set. Then approximate the population mean and the population standard deviation of the data set.

167	180	192	173	145	151	174	175	178	160
195	224	244	146	162	146	177	163	149	188

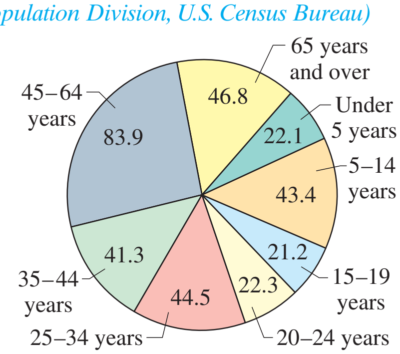
45. Amounts of Caffeine The amounts of caffeine in a sample of five-ounce servings of brewed coffee are shown in the histogram. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set.



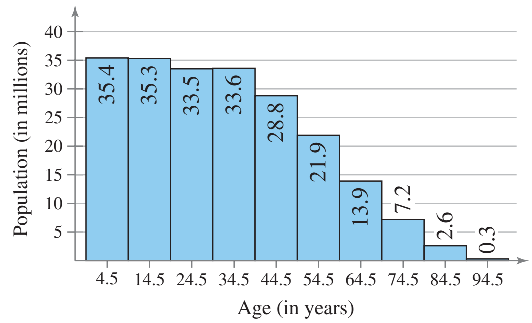
46. Supermarket Trips Thirty people were randomly selected and asked how many trips to the supermarket they had made in the past week. The responses are shown in the histogram. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set.



47. U.S. Population The estimated distribution (in millions) of the U.S. population by age for the year 2015 is shown in the pie chart. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set. Use 70 as the midpoint for "65 years and over." *(Source: Population Division, U.S. Census Bureau)*



48. Brazil's Population Brazil's estimated population for the year 2015 is shown in the histogram. Make a frequency distribution for the data. Then use the table to estimate the sample mean and the sample standard deviation of the data set. *(Adapted from U.S. Census Bureau, International Data Base)*



SC *In Exercises 49 and 50, use StatCrunch to find the sample size, mean, variance, standard deviation, median, range, minimum data value, and maximum data value of the data.*

49. The data represent the total amounts (in dollars) spent by several families at a restaurant.

49 56 75 64 55 49 62 89 30 34 60 52 60 72 75

50. The data represent the prices (in dollars) of several Hewlett-Packard office printers. *(Source: Hewlett-Packard)*

199.99	499.99	149.99	119.99	129.99	229.99
179.99	89.99	299.99	249.99	349.99	99.99

EXTENDING CONCEPTS

51. Coefficient of Variation The **coefficient of variation** CV describes the standard deviation as a percent of the mean. Because it has no units, you can use the coefficient of variation to compare data with different units.

$$CV = \frac{\text{Standard deviation}}{\text{Mean}} \times 100\%$$

The table at the left shows the heights (in inches) and weights (in pounds) of the members of a basketball team. Find the coefficient of variation for each data set. What can you conclude?

Heights	Weights
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

TABLE FOR EXERCISE 51

52. Shortcut Formula You used $SS_x = \Sigma(x - \bar{x})^2$ when calculating variance and standard deviation. An alternative formula that is sometimes more convenient for hand calculations is

$$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n}.$$

You can find the sample variance by dividing the sum of squares by $n - 1$ and the sample standard deviation by finding the square root of the sample variance.

(a) Use the shortcut formula to calculate the sample standard deviations for the data sets given in Exercise 27.

(b) Compare your results with those obtained in Exercise 27.

53. Scaling Data Sample annual salaries (in thousands of dollars) for employees at a company are listed.

42 36 48 51 39 39 42 36 48 33 39 42 45

(a) Find the sample mean and sample standard deviation.

(b) Each employee in the sample is given a 5% raise. Find the sample mean and sample standard deviation for the revised data set.

(c) To calculate the monthly salary, divide each original salary by 12. Find the sample mean and sample standard deviation for the revised data set.

(d) What can you conclude from the results of (a), (b), and (c)?

54. Shifting Data Sample annual salaries (in thousands of dollars) for employees at a company are listed.

40 35 49 53 38 39 40 37 49 34 38 43 47

(a) Find the sample mean and sample standard deviation.

(b) Each employee in the sample is given a $1000 raise. Find the sample mean and sample standard deviation for the revised data set.

(c) Each employee in the sample takes a pay cut of $2000 from their original salary. Find the sample mean and sample standard deviation for the revised data set.

(d) What can you conclude from the results of (a), (b), and (c)?

55. Mean Absolute Deviation Another useful measure of variation for a data set is the **mean absolute deviation** (MAD). It is calculated by the formula

$$\frac{\Sigma|x - \overline{x}|}{n}.$$

(a) Find the mean absolute deviations of the data sets in Exercise 27. Compare your results with the sample standard deviation.

(b) Find the mean absolute deviations of the data sets in Exercise 28. Compare your results with the sample standard deviation.

56. Chebychev's Theorem At least 99% of the data in any data set lie within how many standard deviations of the mean? Explain how you obtained your answer.

57. Pearson's Index of Skewness The English statistician Karl Pearson (1857–1936) introduced a formula for the skewness of a distribution.

$$P = \frac{3(\overline{x} - \text{median})}{s}$$ Pearson's index of skewness

Most distributions have an index of skewness between −3 and 3. When $P > 0$, the data are skewed right. When $P < 0$, the data are skewed left. When $P = 0$, the data are symmetric. Calculate the coefficient of skewness for each distribution. Describe the shape of each.

(a) $\overline{x} = 17, s = 2.3, \text{median} = 19$

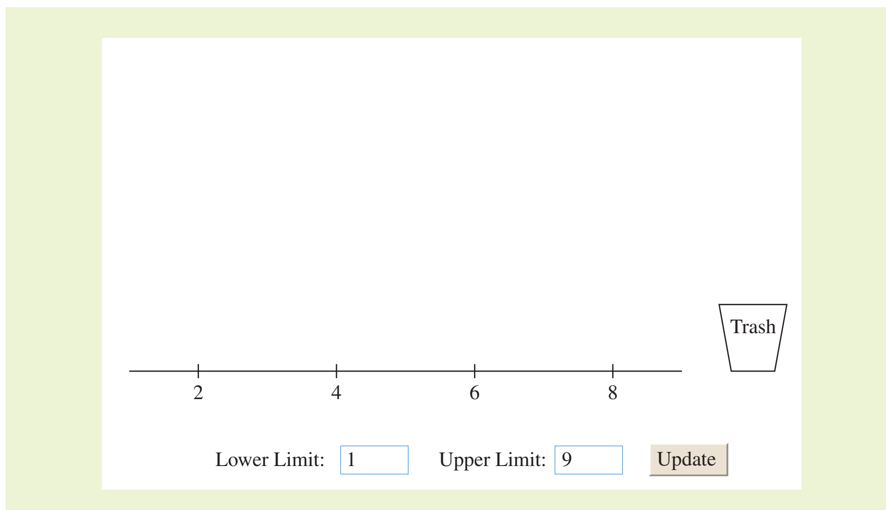
(b) $\overline{x} = 32, s = 5.1, \text{median} = 25$

(c) $\overline{x} = 9.2, s = 1.8, \text{median} = 9.2$

(d) $\overline{x} = 42, s = 6.0, \text{median} = 40$

ACTIVITY 2.4 Standard Deviation

The *standard deviation* applet is designed to allow you to investigate interactively the standard deviation as a measure of spread for a data set. Points can be added to the plot by clicking the mouse above the horizontal axis. The mean of the points is shown as a green arrow. A numeric value for the standard deviation is shown above the plot. Points on the plot can be removed by clicking on the point and then dragging the point into the trash can. All of the points on the plot can be removed by simply clicking inside the trash can. The range of values for the horizontal axis can be specified by inputting lower and upper limits and then clicking UPDATE.



Explore

Step 1 Specify a lower limit.
Step 2 Specify an upper limit.
Step 3 Add 15 points to the plot.
Step 4 Remove all of the points from the plot.

Draw Conclusions

1. Specify the lower limit to be 10 and the upper limit to be 20. Plot 10 points that have a mean of about 15 and a standard deviation of about 3. Write the estimates of the values of the points. Plot a point with a value of 15. What happens to the mean and standard deviation? Plot a point with a value of 20. What happens to the mean and standard deviation?

2. Specify the lower limit to be 30 and the upper limit to be 40. How can you plot eight points so that the points have the largest possible standard deviation? Use the applet to plot the set of points and then use the formula for standard deviation to confirm the value given in the applet. How can you plot eight points so that the points have the lowest possible standard deviation? Explain.

Earnings of Athletes

The earnings of professional athletes in different sports can vary. An athlete can be paid a base salary, earn signing bonuses upon signing a new contract, or even earn money by finishing in a certain position in a race or tournament. The data shown below are the earnings (for performance only, no endorsements) from Major League Baseball (MLB), Major League Soccer (MLS), the National Basketball Association (NBA), the National Football League (NFL), the National Hockey League (NHL), the National Association for Stock Car Auto Racing (NASCAR), and the Professional Golf Association Tour (PGA) for a recent year.

Organization	Number of players
MLB	858
MLS	410
NBA	463
NFL	1861
NHL	722
NASCAR	76
PGA	262

Number of Players Separated into Earnings Ranges

Organization	\$0–\$500,000	\$500,001–\$2,000,000	\$2,000,001–\$6,000,000	\$6,000,001–\$10,000,000	\$10,000,001+
MLB	353	182	164	85	74
MLS	403	5	1	1	0
NBA	35	157	137	77	57
NFL	554	746	438	85	38
NHL	42	406	237	37	0
NASCAR	23	16	31	6	0
PGA	110	115	36	1	0

EXERCISES

1. **Revenue** Which organization had the greatest total player earnings? Explain your reasoning.
2. **Mean Earnings** Estimate the mean earnings of a player in each organization. Use \$19,000,000 as the midpoint for \$10,000,001+.
3. **Revenue** Which organization had the greatest earnings per player? Explain your reasoning.
4. **Standard Deviation** Estimate the standard deviation for the earnings of a player in each organization. Use \$19,000,000 as the midpoint for \$10,000,001+.
5. **Standard Deviation** Which organization had the greatest standard deviation? Explain your reasoning.
6. **Bell-Shaped Distribution** Of the seven organizations, which is most bell-shaped? Explain your reasoning.

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 SC Report 14

Finding the Quartiles of a Data Set

The number of nuclear power plants in the top 15 nuclear power-producing countries in the world are listed. Find the first, second, and third quartiles of the data set. What can you conclude? *(Source: International Atomic Energy Agency)*

7 18 11 6 59 17 18 54 104 20 31 8 10 15 19

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: 6 7 8 10 11 15 17

18

Upper half: 18 19 20 31 54 59 104

Q_1 = 10, Q_2 = 18, Q_3 = 31

Interpretation About one fourth of the countries have 10 or fewer nuclear power plants; about one half have 18 or fewer; and about three fourths have 31 or fewer.

Try It Yourself 1

Find the first, second, and third quartiles for the ages of the 50 richest people using the data set listed in the Chapter Opener on page 37. What can you conclude?

a. *Order* the data set.
b. Find the *median* Q_2.
c. Find the *first and third quartiles*, Q_1 and Q_3.
d. *Interpret* the results in the context of the data.

Answer: Page A34

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. What can you conclude?

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 Plus 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: Tuition

Variable	N	Mean	SE Mean	StDev	Minimum
Tuition	25	23.960	0.788	3.942	15.000

Variable	Q1	Median	Q3	Maximum
Tuition	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 PLUS

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. What can you conclude?

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

Answer: Page A34

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

DEFINITION

The **interquartile range (IQR)** of a data set is a measure of variation that gives the range of the middle 50% of the data. It is the difference between the third and first quartiles.

$$\text{Interquartile range (IQR)} = Q_3 - Q_1$$

EXAMPLE 3

Finding the Interquartile Range

Find the interquartile range of the data set given in Example 1. What can you conclude from the result?

Solution

From Example 1, you know that $Q_1 = 10$ and $Q_3 = 31$. So, the interquartile range is

$$\text{IQR} = Q_3 - Q_1 = 31 - 10 = 21.$$

Interpretation The number of power plants in the middle portion of the data set vary by at most 21.

Try It Yourself 3

Find the interquartile range for the ages of the 50 richest people listed in the Chapter Opener on page 37.

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 A34

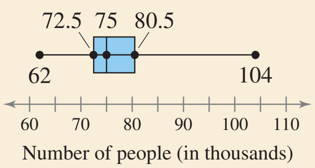
The IQR can also be used to identify outliers. First, multiply the IQR by 1.5. Then subtract that value from Q_1, and add that value to Q_3. Any data value that is smaller than $Q_1 - 1.5(\text{IQR})$ or larger than $Q_3 + 1.5(\text{IQR})$ is an outlier. For instance, the IQR in Example 1 is $31 - 10 = 21$ and $1.5(21) = 31.5$. So, adding 31.5 to Q_3 gives $Q_3 + 31.5 = 31 + 31.5 = 62.5$. Because $104 > 62.5$, 104 is an outlier.

Another important application of quartiles is to represent data sets using box-and-whisker plots. A **box-and-whisker plot** (or **boxplot**) 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 44 Super Bowls played, Super Bowl XIV had the highest attendance at about 104,000. Super Bowl I had the lowest attendance at about 62,000. The box-and-whisker plot summarizes the attendances (in thousands of people) at the 44 Super Bowls. (Source: National Football League)

Super Bowl Attendance



About how many Super Bowl attendances are represented by the right whisker? About how many are represented by the left whisker?

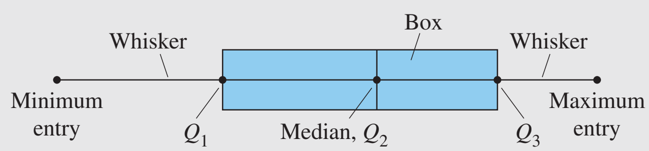
1. The minimum entry
2. The first quartile Q_1
3. The median Q_2
4. The third quartile Q_3
5. The maximum entry

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.



EXAMPLE 4

SC Report 15

▸ Drawing a Box-and-Whisker Plot

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

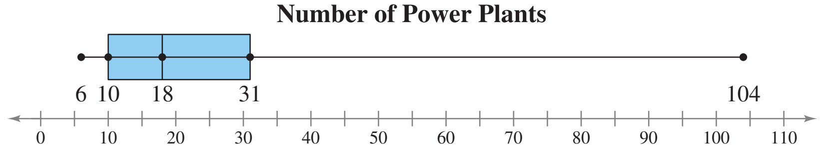
See MINITAB and TI-83/84 Plus steps on pages 122 and 123.

▸ Solution

The five-number summary of the data set is displayed below. Using these five numbers, you can construct the box-and-whisker plot shown.

Min = 6, $Q_1 = 10$, $Q_2 = 18$, $Q_3 = 31$, Max = 104,

Number of Power Plants



Interpretation You can make several conclusions from the display. One is that about half the data values are between 10 and 31. By looking at the length of the right whisker, you can also conclude that the data value of 104 is a possible outlier.

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.

▸ Try It Yourself 4

Draw a box-and-whisker plot that represents the ages of the 50 richest people listed in the Chapter Opener on page 37. What can you conclude?

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 A34

INSIGHT

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

PERCENTILES AND OTHER FRACTILES

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.

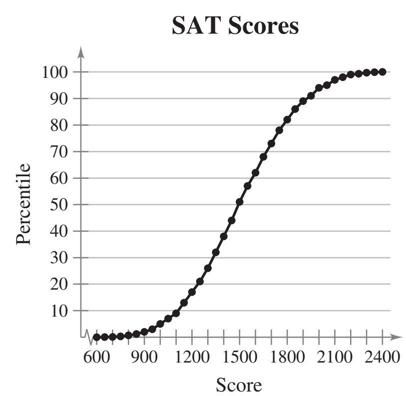
Fractiles	Summary	Symbols
Quartiles	Divide a data set into 4 equal parts.	Q_1, Q_2, Q_3
Deciles	Divide a data set into 10 equal parts.	$D_1, D_2, D_3, \ldots, D_9$
Percentiles	Divide a data set into 100 equal parts.	$P_1, P_2, P_3, \ldots, P_{99}$

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.

EXAMPLE 5

Interpreting Percentiles

The ogive at the right represents the cumulative frequency distribution for SAT test scores of college-bound students in a recent year. What test score represents the 62nd percentile? How should you interpret this? *(Source: The College Board)*



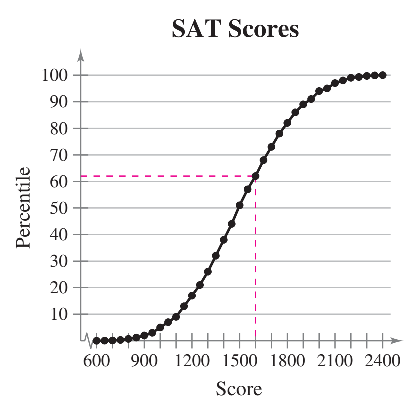
STUDY TIP

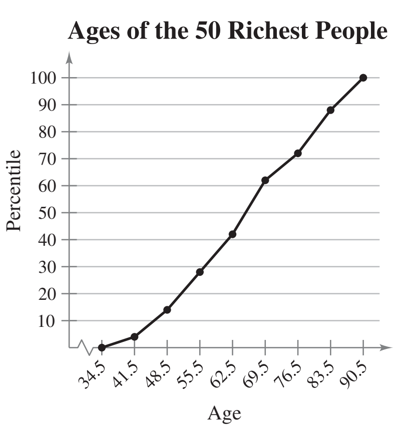
It is important that you understand what a percentile means. For instance, if the weight of a six-month-old infant is at the 78th percentile, the infant weighs more than 78% of all six-month-old infants. It does not mean that the infant weighs 78% of some ideal weight.

Solution

From the ogive, you can see that the 62nd percentile corresponds to a test score of 1600.

Interpretation This means that approximately 62% of the students had an SAT score of 1600 or less.





Try It Yourself 5

The ages of the 50 richest people are represented in the cumulative frequency graph at the left. At what percentile is someone who is 66 years old? How should you interpret this?

a. *Use the graph* to find the percentile that corresponds to the given age.
b. *Interpret* the results in the context of the data. *Answer: Page A34*

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 less than the mean. If z is positive, the corresponding x-value is greater than 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 speeds 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 $\qquad$ $x = 47$ mph $\qquad$ $x = 56$ mph

$$z = \frac{62 - 56}{4} = 1.5 \qquad z = \frac{47 - 56}{4} = -2.25 \qquad z = \frac{56 - 56}{4} = 0$$

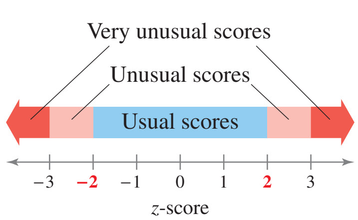
Interpretation From the z-scores, you can conclude that a speed of 62 miles per hour is 1.5 standard deviations above the mean; a speed of 47 miles per hour is 2.25 standard deviations below the mean; and a speed of 56 miles per hour is equal to the mean. If the distribution of the speeds is approximately 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. *Transform* each value to a z-score.
b. *Interpret* the results.

Answer: Page A34



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 2009, Heath Ledger won the Oscar for Best Supporting Actor at age 29 for his role in the movie *The Dark Knight*. Penelope Cruz won the Oscar for Best Supporting Actress at age 34 for her role in *Vicky Cristina Barcelona*. The mean age of all Best Supporting Actor winners is 49.5, with a standard deviation of 13.8. The mean age of all Best Supporting Actress winners is 39.9, with a standard deviation of 14.0. Find the z-scores that correspond to the ages of Ledger and Cruz. Then compare your results.

▶ Solution

The z-scores that correspond to the ages of the two performers are calculated below.

Heath Ledger

$$z = \frac{x - \mu}{\sigma} = \frac{29 - 49.5}{13.8} \approx -1.49$$

Penelope Cruz

$$z = \frac{x - \mu}{\sigma} = \frac{34 - 39.9}{14.0} \approx -0.42$$

The age of Heath Ledger was 1.49 standard deviations below the mean, and the age of Penelope Cruz was 0.42 standard deviation below the mean.

Interpretation Compared with other Best Supporting Actor winners, Heath Ledger was relatively younger, whereas the age of Penelope Cruz was only slightly lower than the average age of other Best Supporting Actress winners. Both z-scores fall between -2 and 2, so neither score would be considered unusual.

▶ Try It Yourself 7

In 2009, Sean Penn won the Oscar for Best Actor at age 48 for his role in the movie *Milk*. Kate Winslet won the Oscar for Best Actress at age 33 for her role in *The Reader*. The mean age of all Best Actor winners is 43.7, with a standard deviation of 8.7. The mean age of all Best Actress winners is 35.9, with a standard deviation of 11.4. Find the z-scores that correspond to the ages of Penn and Winslet. Then compare your results.

a. *Identify* μ and σ for each data set.
b. *Transform* each value to a z-score.
c. *Compare* your results.

Answer: Page A34

2.5 EXERCISES

BUILDING BASIC SKILLS AND VOCABULARY

1. 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?

2. 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?

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

4. 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?

5. Explain how the interquartile range of a data set can be used to identify outliers.

6. Describe the relationship between quartiles and percentiles.

True or False? *In Exercises 7–14, determine whether the statement is true or false. If it is false, rewrite it as a true statement.*

7. The mean and median of a data set are both fractiles.

8. About one quarter of a data set falls below Q_1.

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

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

11. The 50th percentile is equivalent to Q_1.

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

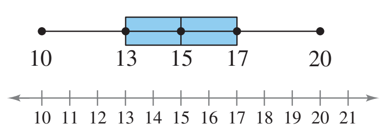
13. A z-score of -2.5 is considered very unusual.

14. A z-score of 1.99 is considered usual.

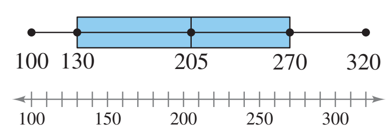
USING AND INTERPRETING CONCEPTS

Graphical Analysis *In Exercises 15–20, use the box-and-whisker plot to identify (a) the five-number summary, and (b) the interquartile range.*

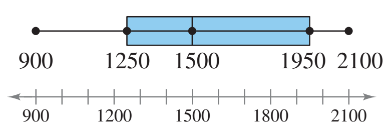
15.



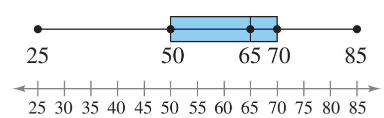
16.



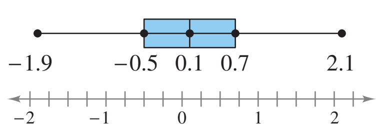
17.



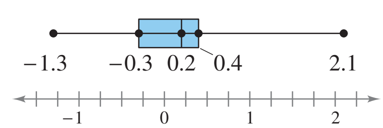
18.



19.



20.



In Exercises 21–24, (a) find the five-number summary, and (b) draw a box-and-whisker plot of the data.

21. 39 36 30 27 26 24 28 35 39 60 50 41 35 32 51

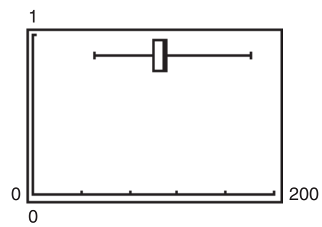
22. 171 176 182 150 178 180 173 170 174 178 181 180

23. 4 7 7 5 2 9 7 6 8 5 8 4 1 5 2 8 7 6 6 9

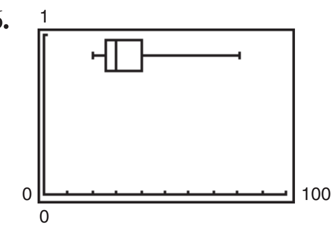
24. 2 7 1 3 1 2 8 9 9 2 5 4 7 3 7 5 4 7
2 3 5 9 5 6 3 9 3 4 9 8 8 2 3 9 5

Interpreting Graphs *In Exercises 25–28, 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.*

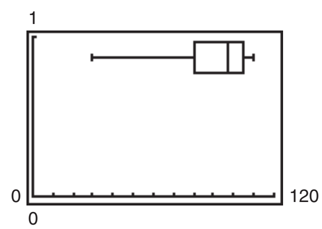
25.



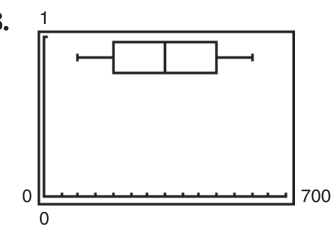
26.



27.



28.



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

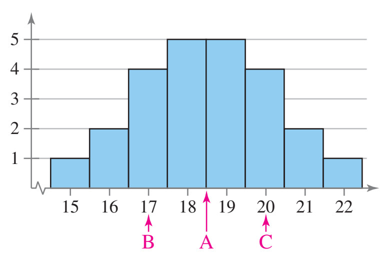


FIGURE FOR EXERCISE 29

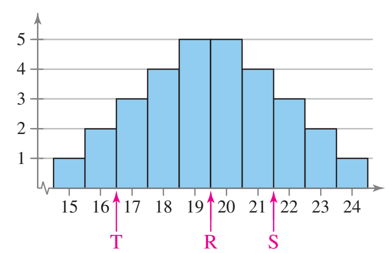


FIGURE FOR EXERCISE 30

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

Using Technology to Find Quartiles and Draw Graphs *In Exercises 31–34, 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.*

31. 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

32. 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

33. 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		

34. 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		

35. TV Viewing Refer to the data set given in Exercise 31 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.

36. Manufacturer Earnings Refer to the data set given in Exercise 34 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.

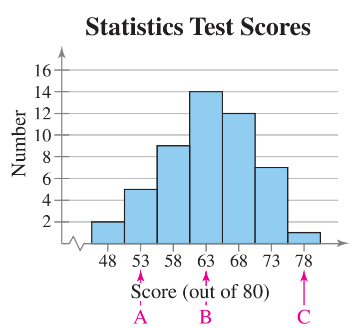
Graphical Analysis

In Exercises 37 and 38, the midpoints A, B, and C are marked on the histogram. Match them with the indicated z-scores. Which z-scores, if any, would be considered unusual?

37. $z = 0$

$z = 2.14$

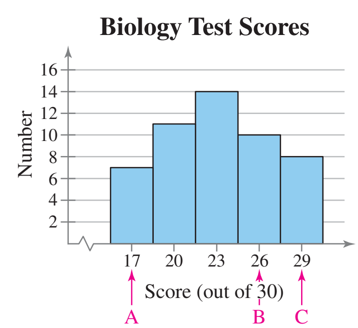
$z = -1.43$



38. $z = 0.77$

$z = 1.54$

$z = -1.54$



Comparing Test Scores *For the statistics test scores in Exercise 37, the mean is 63 and the standard deviation is 7.0, and for the biology test scores in Exercise 38, the mean is 23 and the standard deviation is 3.9. In Exercises 39–42, 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.

39. A student gets a 75 on the statistics test and a 25 on the biology test.

40. A student gets a 60 on the statistics test and a 22 on the biology test.

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

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

43. Life Spans of Tires A certain brand of automobile tire has a mean life span of 35,000 miles, with 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 30,000 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.

44. 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 45–50, 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 86) with a mean of 69.9 inches and a standard deviation of 3.0 inches. (Adapted from National Center for Health Statistics)*

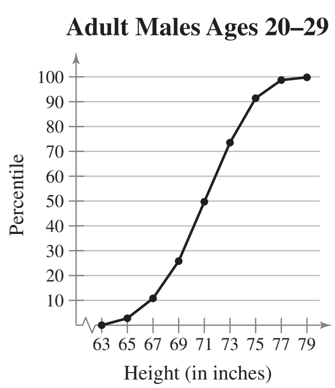


FIGURE FOR EXERCISES 45–50

45. What height represents the 60th percentile? How should you interpret this?

46. What percentile is a height of 77 inches? How should you interpret this?

47. 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.

48. 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.

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

50. 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

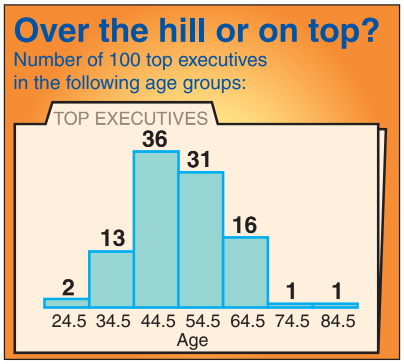


FIGURE FOR EXERCISE 51

51. 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) Find the five-number summary.

(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 52–55, find the midquartile of the given data set.

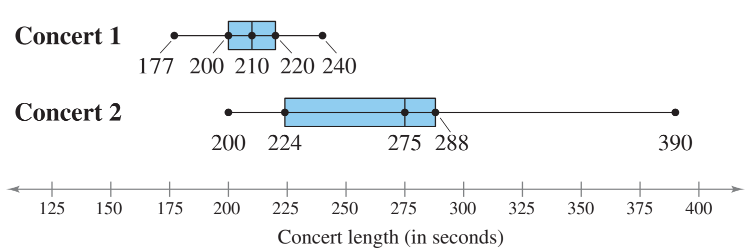
52. 5 7 1 2 3 10 8 7 5 3

53. 23 36 47 33 34 40 39 24 32 22 38 41

54. 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

55. 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

56. 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 lengths (in seconds) of songs played at two different concerts are shown.



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

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

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

(d) Can you determine which concert lasted longer? Explain.

57. 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.*

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

In Exercises 58 and 59, use the information from Example 7 and the fact that there have been 73 Oscars for Best Supporting Actor and 73 Oscars for Best Supporting Actress awarded.

58. Only three winners were younger than Heath Ledger when they won the Oscar for Best Supporting Actor. Find the percentile that corresponds to Heath Ledger's age.

59. Forty-three winners were older than Penelope Cruz when they won the Oscar for Best Supporting Actress. Find the percentile that corresponds to Penelope Cruz's age.

Modified Boxplot *A* ***modified boxplot*** *is a boxplot that uses symbols to identify outliers. The horizontal line of a modified boxplot extends as far as the minimum data value that is not an outlier and the maximum data value that is not an outlier. In Exercises 60 and 61, (a) identify any outliers (using the 1.5 × IQR rule), and (b) draw a modified boxplot that represents the data set. Use asterisks (*) to identify outliers.*

60. 16 9 11 12 8 10 12 13 11 10 24 9 2 15 7

61. 75 78 80 75 62 72 74 75 80 95 76 72

SC *In Exercises 62 and 63, use StatCrunch to (a) find the five-number summary, (b) construct a regular boxplot, and (c) construct a modified boxplot for the data.*

62. The data represent the speeds (in miles per hour) of several vehicles.

68 88 70 72 70 69 72 62 65 70 75 52 65

63. The data represent the weights (in pounds) of several professional football players.

225	250	305	285	275	265	290	310	290	250	210	225
308	325	260	165	195	245	235	298	395	255	268	190

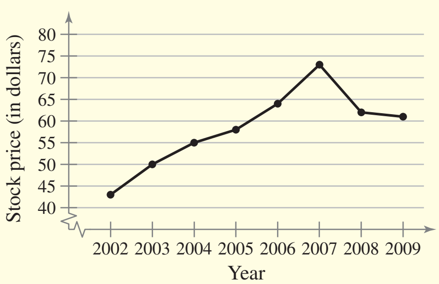
USES AND ABUSES

Statistics in the Real World

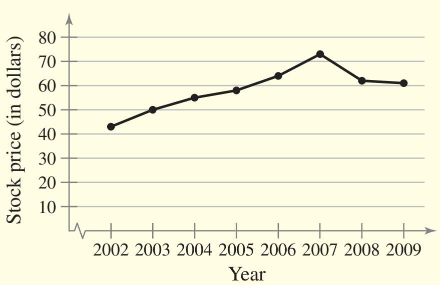
Uses

Descriptive statistics help you see trends or patterns in 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.

Procter & Gamble's Stock Price



Procter & Gamble's Stock Price



Abuses

Knowing how statistics are calculated can help you analyze questionable statistics. For instance, 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 the company used the mean because it is a greater amount.

Statistical graphs can also be misleading. Compare the two time series charts at the left, which show the year-end stock prices for the Procter & Gamble Corporation. The data are the same for each chart. The first graph, however, has a cropped vertical axis, which makes it appear that the stock price increased greatly from 2002 to 2007, then decreased greatly from 2007 to 2009. In the second graph, the scale on the vertical axis begins at zero. This graph correctly shows that the stock price changed modestly during this time period. *(Source: Procter & Gamble Corporation)*

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. Use the Internet or some other resource to find an example of a graph that might lead to incorrect conclusions.

2. You are publishing an article that discusses how eating oatmeal can help lower cholesterol. Because eating oatmeal might help people with high cholesterol, you include a graph that exaggerates the effects of eating oatmeal on lowering cholesterol. Do you think it is ethical to publish this graph? Explain.

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 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	*1–6*	*13, 14*
■ How to find a weighted mean of a data set and the mean of a frequency distribution	*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	*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	*9, 10*	*35, 36*
Section 2.5		
■ How to find the quartiles and interquartile range of a data set	*1–3*	*37, 38, 41*
■ How to draw a box-and-whisker plot	*4*	*39, 40, 42*
■ How to interpret other fractiles such as percentiles	*5*	*43, 44*
■ How to find and interpret the standard score (z-score)	*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 number of students per faculty member for 20 public colleges. *(Source: Kiplinger)*

13	15	15	8	16	20	28	19	18	15
21	23	30	17	10	16	15	16	20	15

1. Make a frequency distribution of the data set using five classes. Include the class limits, midpoints, 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 of the data set 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 of the data set 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 air quality indices for 30 U.S. cities. *(Source: AIRNow)*

25	35	20	75	10	10	61	89	44	22
34	33	38	30	47	53	44	57	71	20
42	52	48	41	35	59	53	61	65	25

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 Houston. Use the data to construct a scatter plot. What type of pattern is shown in the scatter plot? *(Source: Emporis Corporation)*

Height (in feet)	992	780	762	756	741	732	714	662	579
Number of stories	71	56	53	55	47	53	50	49	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	1998	1999	2000	2001	2002	2003
Unemployment rate	4.5	4.2	4.0	4.7	5.8	6.0

Year	2004	2005	2006	2007	2008	2009
Unemployment rate	5.5	5.1	4.6	4.6	5.8	9.3

In Exercises 11 and 12, use the following data set. The data set represents the results of a survey that asked U.S. adults where they would be at midnight when the new year arrived. (Adapted from Rasmussen Reports)

Response	At home	At friend's home	At restaurant or bar	Somewhere else	Not sure
Number	620	110	50	100	130

11. Make a Pareto chart of the data set.

12. Make a pie chart of the data set.

SECTION 2.3

In Exercises 13 and 14, find the mean, median, and mode of the data, if possible. If any of these measures cannot be found or a measure does not represent the center of the data, explain why.

13. Vertical Jumps The vertical jumps (in inches) of a sample of 10 college basketball players at the 2009 NBA Draft Combine *(Source: Sports Phenoms, Inc.)*

26.0 29.5 27.0 30.5 29.5 25.0 31.5 33.0 32.0 27.5

14. Airport Scanners The responses of 542 adults who were asked whether they approved the use of full-body scanners at airport security checkpoints *(Adapted from USA Today/Gallup Poll)*

Approved: 423 Did not approve: 108 No opinion: 11

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.

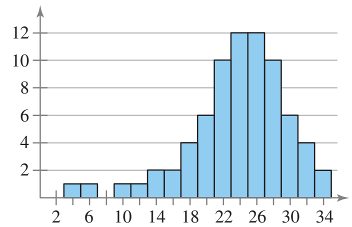
96 85 91 86

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

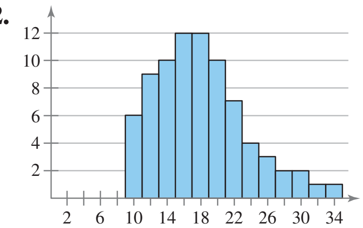
20. Describe the shape of the distribution in the histogram you made in Exercise 4. Is the distribution symmetric, uniform, or skewed?

In Exercises 21 and 22, determine whether the approximate shape of the distribution in the histogram is symmetric, uniform, skewed left, skewed right, or none of these. Justify your answer.

21.



22.



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

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

SECTION 2.4

25. The data set represents the mean prices of movie tickets (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 prices of movie tickets (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 mileages (in thousands of miles) for a rental car company's fleet are listed. Find the population mean and the population standard deviation of the data.

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

28. The ages of the Supreme Court justices as of January 27, 2010 are listed. Find the population mean and the population standard deviation of the data. *(Source: Supreme Court of the United States)*

55 89 73 73 61 76 71 59 55

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 the sample 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 for 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 the data set has a bell-shaped distribution.)

32. The mean rate for satellite television for 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 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. Using Chebychev's Theorem, determine 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. Using Chebychev's Theorem, determine 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 televisions are listed. Find the sample mean and the sample 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 the sample 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 fuel economies (in highway miles per gallon) of several Harley-Davidson motorcycles. (Source: Total Motorcycle)

53 57 60 57 54 53 54 53 54 42 48
53 47 47 50 48 42 42 54 54 60

37. Find the five-number summary of the data set.

38. Find the interquartile range.

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

40. About how many motorcycles fall on or below the third quartile?

41. Find the interquartile range of the data from Exercise 13.

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

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 75 represents the 65th percentile of the grades. What percent of students scored higher than 75?

44. As of January 2010, there were 755 "oldies" radio stations in the United States. If one station finds that 104 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 towing capacities (in pounds) of 25 four-wheel drive pickup trucks have a bell-shaped distribution, with a mean of 11,830 pounds and a standard deviation of 2370 pounds. Use z-scores to determine if the towing capacities of the following randomly selected four-wheel drive pickup trucks are unusual.

45. 16,500 pounds

46. 5500 pounds

47. 18,000 pounds

48. 11,300 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 represents 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, boundaries, frequencies, 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 stem-and-leaf plot. Use one line per stem.

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

(g) Display the data using an ogive.

2. Use frequency distribution formulas to approximate the sample mean and the sample 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 (10.6), footwear (17.2), equipment (24.9), and recreational transport (27.0). 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, median, and 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 regular season wins for each Major League Baseball team in 2009 are listed. *(Source: Major League Baseball)*

103	95	84	75	64	87	86	79	65	65	97	87	85	75	93
87	86	70	59	91	83	80	78	74	62	95	92	88	75	70

(a) Find the five-number summary of the data set.

(b) Find the interquartile range.

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

PUTTING IT ALL TOGETHER

Real Statistics — Real Decisions

You are a member of your local apartment association. The association represents rental housing owners and managers who operate residential rental property throughout the greater metropolitan area. Recently, the association has received several complaints from tenants in a particular area of the city who feel that their monthly rental fees are much higher compared to other parts of the city.

You want to investigate the rental fees. You gather the data shown in the table at the right. Area A represents the area of the city where tenants are unhappy about their monthly rents. The data represent the monthly rents paid by a random sample of tenants in Area A and three other areas of similar size. Assume all the apartments represented are approximately the same size with the same amenities.

NATIONAL APARTMENT ASSOCIATION

AMERICA'S LEADING ADVOCATE FOR QUALITY RENTAL HOUSING

The Monthly Rents (in dollars) Paid by 12 Randomly Selected Apartment Tenants in 4 Areas of Your City

Area A	Area B	Area C	Area D
1275	1124	1085	928
1110	954	827	1096
975	815	793	862
862	1078	1170	735
1040	843	919	798
997	745	943	812
1119	796	756	1232
908	816	765	1036
890	938	809	998
1055	1082	1020	914
860	750	710	1005
975	703	775	930

EXERCISES

1. *How Would You Do It?*

(a) How would you investigate the complaints from renters who are unhappy about their monthly rents?

(b) Which statistical measure do you think would best represent the data sets for the four areas of the city?

(c) Calculate the measure from part (b) for each of the four areas.

2. *Displaying the Data*

(a) What type of graph would you choose to display the data? Explain your reasoning.

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

(c) Based on your data displays, does it appear that the monthly rents in Area A are higher than the rents in the other areas of the city? Explain.

3. *Measuring the Data*

(a) What other statistical measures in this chapter could you use to analyze the monthly rent data?

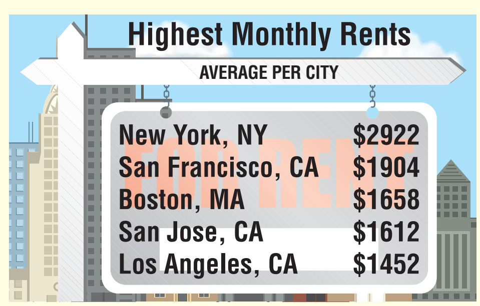
(b) Calculate the measures from part (a).

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

4. *Discussing the Data*

(a) Do you think the complaints in Area A are legitimate? How do you think they should be addressed?

(b) What reasons might you give as to why the rents vary among different areas of the city?



(Source: Forbes)

TECHNOLOGY

MINITAB EXCEL TI-83/84 PLUS

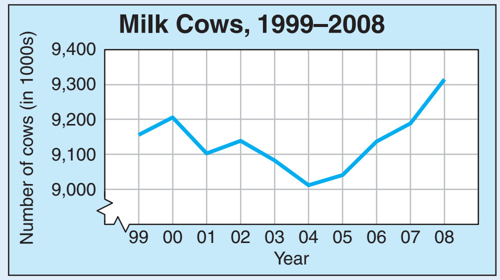
Dairy Farmers of America is an association that provides help to dairy farmers. Part of this help is gathering and distributing statistics on milk production.

MONTHLY MILK PRODUCTION

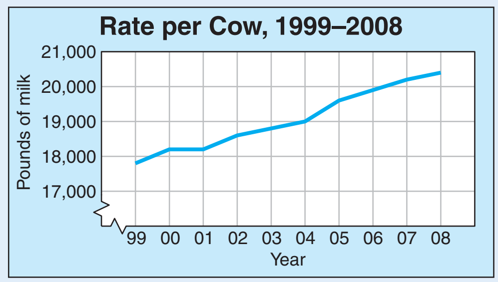
The following data set was supplied by a dairy farmer. It lists the monthly milk productions (in pounds) for 50 Holstein dairy cows. *(Source: Matlink Dairy, Clymer, NY)*

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

www.dfamilk.com



(Source: National Agricultural Statistics Service)



(Source: National Agricultural Statistics Service)

From 1999 to 2008, the number of dairy cows in the United States increased by only 1.7% while the yearly milk production per cow increased by almost 15%.

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.

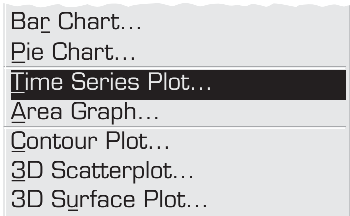
6. Use the frequency distribution to estimate the sample mean of the data. Compare your results with Exercise 1.
7. Use the frequency distribution to find the sample standard deviation for the data. Compare your results with Exercise 2.
8. **Writing** Use the results of Exercises 6 and 7 to write a general 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 Plus.

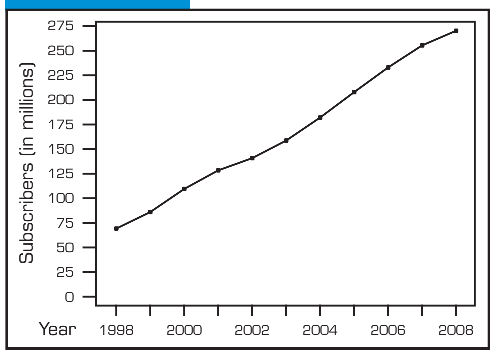
2 USING TECHNOLOGY TO DETERMINE DESCRIPTIVE STATISTICS

Here are some MINITAB and TI-83/84 Plus printouts for three examples in this chapter.

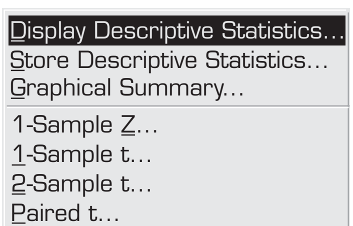
(See Example 7, page 59.)



MINITAB



(See Example 4, page 83.)



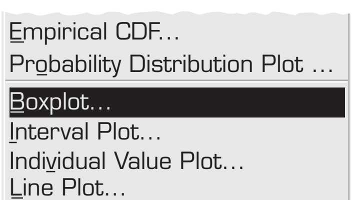
MINITAB

Descriptive Statistics: Salaries

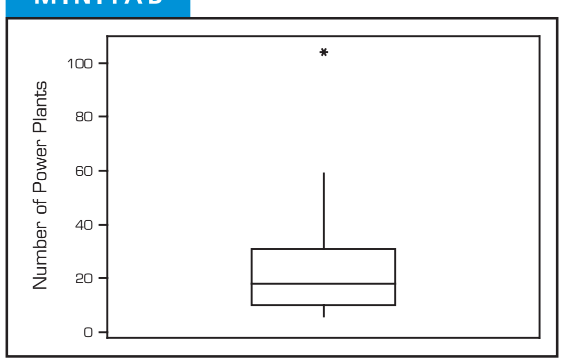
Variable	N	Mean	SE Mean	StDev	Minimum
Salaries	10	41.500	0.992	3.136	37.000

Variable	Q1	Median	Q3	Maximum
Salaries	38.750	41.000	44.250	47.000

(See Example 4, page 103.)



MINITAB



(See Example 7, page 59.)

TI-83/84 PLUS

STAT PLOTS
1: Plot1...Off
L1 L2 □
2: Plot2...Off
L1 L2 □
3: Plot3...Off
L1 L2 □

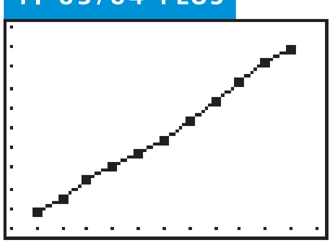
TI-83/84 PLUS

Plot1 Plot2 Plot3
On Off
Type:
Xlist: L1
Ylist: L2
Mark: □ + .

TI-83/84 PLUS

ZOOM MEMORY
4↑ ZDecimal
5: ZSquare
6: ZStandard
7: ZTrig
8: ZInteger
9: ZoomStat
0: ZoomFit

TI-83/84 PLUS



(See Example 4, page 83.)

TI-83/84 PLUS

EDIT CALC TESTS
1: 1-Var Stats
2: 2-Var Stats
3: Med-Med
4: LinReg(ax+b)
5: QuadReg
6: CubicReg
7↓ QuartReg

TI-83/84 PLUS

1-Var Stats L1

TI-83/84 PLUS

1-Var Stats
$\bar{x}$= 41.5
Σx= 415
Σx^2= 17311
Sx= 3.13581462
σx= 2.974894956
↓n= 10

(See Example 4, page 103.)

TI-83/84 PLUS

STAT PLOTS
1: Plot1...Off
L1 L2 □
2: Plot2...Off
L1 L2 □
3: Plot3...Off
L1 L2 □
4↓ PlotsOff

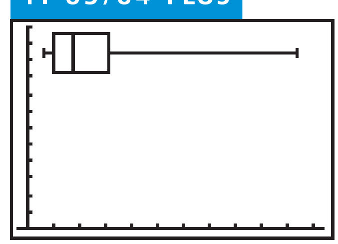
TI-83/84 PLUS

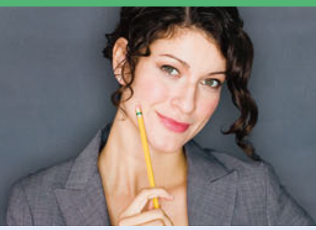
Plot1 Plot2 Plot3
On Off
Type:
Xlist: L1
Freq: 1

TI-83/84 PLUS

ZOOM MEMORY
4↑ ZDecimal
5: ZSquare
6: ZStandard
7: ZTrig
8: ZInteger
9: ZoomStat
0: ZoomFit

TI-83/84 PLUS





CUMULATIVE REVIEW

Chapters 1 and 2

In Exercises 1 and 2, identify the sampling technique used and discuss potential sources of bias (if any). Explain.

1. For quality assurance, every fortieth toothbrush is taken from each of four assembly lines and tested to make sure the bristles stay in the toothbrush.

2. Using random digit dialing, researchers asked 1200 U.S. adults their thoughts on health care reform.

3. In 2008, a worldwide study of all airlines found that baggage delays were caused by transfer baggage mishandling (49%), failure to load at originating airport (16%), arrival station mishandling (8%), space-weight restriction (6%), loading/offloading error (5%), tagging error (3%), and ticketing error/bag switch/security/other (13%). Use a Pareto chart to organize the data. *(Source: Société International de Télécommunications Aéronautiques)*

In Exercises 4 and 5, determine whether the numerical value is a parameter or a statistic. Explain your reasoning.

4. In 2009, the average salary of a Major League Baseball player was \$2,996,106. *(Source: Major League Baseball)*

5. In a recent survey of 1000 voters, 19% said that First Lady of the United States Michelle Obama will be very involved in policy decisions. *(Source: Rasmussen Reports)*

6. The mean annual salary for a sample of electrical engineers is \$83,500, with a standard deviation of \$1500. The data set has a bell-shaped distribution.

(a) Use the Empirical Rule to estimate the number of electrical engineers whose annual salaries are between \$80,500 and \$86,500.

(b) If 40 additional electrical engineers were sampled, about how many of these electrical engineers would you expect to have annual salaries between \$80,500 and \$86,500?

(c) The salaries of three randomly selected electrical engineers are \$90,500, \$79,750, and \$82,600. Find the z-score that corresponds to each salary. According to the z-scores, would the salaries of any of these engineers be considered unusual?

In Exercises 7 and 8, identify the population and the sample.

7. A survey of career counselors at 195 colleges and universities found that 90% of the students working with their offices were interested in federal jobs or internships. *(Source: Partnership for Public Service Survey)*

8. A study of 232,606 people was conducted to find a link between taking antioxidant vitamins and living a longer life. *(Source: Journal of the American Medical Association)*

In Exercises 9 and 10, decide which method of data collection you would use to collect data for the study. Explain.

9. A study of the years of service of the 100 members of the Senate

10. A study of the effects of removing recess from schools

In Exercises 11 and 12, determine whether the data are qualitative or quantitative and identify the data set's level of measurement.

11. The number of games started by pitchers with at least one start for the New York Yankees in 2009 are listed. *(Source: Major League Baseball)*

9 34 1 33 32 31 7 9 6

12. The five top-earning states in 2008 by median income are listed. *(Source: U.S. Census Bureau)*

1. Maryland 2. New Jersey 3. Connecticut 4. Alaska 5. Hawaii

13. The number of tornadoes by state in a recent year is listed. (a) Find the data set's five-number summary, (b) draw a box-and-whisker plot that represents the data set, and (c) describe the shape of the distribution. *(Source: National Climatic Data Center)*

81	1	8	69	30	34	0	0	56	54
2	6	21	14	46	136	17	23	2	0
1	5	71	105	39	10	40	1	0	7
4	0	23	53	4	27	1	11	0	14
19	23	105	4	0	24	4	0	63	6

14. Five test scores are given. The first four test scores are 15% of the final grade, and the last test score is 40% of the final grade. Find the weighted mean of the test scores.

85 92 84 89 91

15. Tail lengths (in feet) for a sample of American alligators are listed.

6.5 3.4 4.2 7.1 5.4 6.8 7.5 3.9 4.6

(a) Find the mean, median, and mode of the tail lengths. Which best describes a typical American alligator tail length? Explain your reasoning.

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

16. A study shows that the number of deaths due to heart disease for women has decreased every year for the past five years.

(a) Make an inference based on the results of the study.

(b) What is wrong with this type of reasoning?

In Exercises 17–19, use the following data set. The data represent the points scored by each player on the Montreal Canadiens in a recent NHL season. (Source: National Hockey League)

5	64	50	1	41	0	39	23	32	28
26	23	33	23	22	1	17	18	12	11
11	9	65	3	2	41	21	1	0	39

17. Make a frequency distribution using eight classes. Include the class limits, midpoints, boundaries, frequencies, relative frequencies, and cumulative frequencies.

18. Describe the shape of the distribution.

19. Make a relative frequency histogram using the frequency distribution in Exercise 17. Then determine which class has the greatest relative frequency and which has the least relative frequency.