

# STATISTICS AND DATA ANALYSIS

## CHAPTER OBJECTIVES

- Make and use bar graphs, histograms, frequency distribution tables, stem-and-leaf plots, and box-and-whisker plots. (*Lessons 14-1, 14-2, 14-3*)
- Find the measures of central tendency and the measures of variability. (*Lessons 14-2, 14-3*)
- Use the normal distribution curve. (*Lesson 14-4*)
- Find the standard error of the mean to predict the true mean of a population with a certain level of confidence. (*Lesson 14-5*)

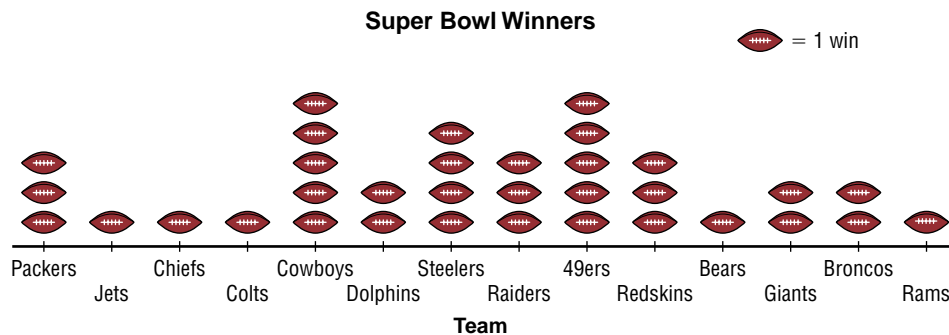
# The Frequency Distribution

## OBJECTIVES

- Draw, analyze, and use bar graphs and histograms.
- Organize data into a frequency distribution table.



**FOOTBALL** The AFL-NFL World Championship Game, as it was originally called, became the Super Bowl in 1969. The graph below shows the first 34 Super Bowl winners. What team has won the most Super Bowls?



By looking at the graph, you can quickly determine that the Dallas Cowboys and the San Francisco 49ers have both won five Super Bowls.

A graph is often used to provide a picture of statistical data. One advantage of using a graph to show data is that a person can easily see any relationships or patterns that may exist. The number of Super Bowl victories for various teams is depicted as a **line plot**. A line plot uses symbols to show frequency. A **bar graph** can show the same information by using bars to indicate the frequency.

A **back-to-back bar graph** is a special bar graph that shows the comparisons of two sets of related data. A back-to-back graph is plotted on a coordinate system with the horizontal scale repeated in each direction from the central axis.

## Example



**1 ECONOMICS** The following data relates the amount of education with the median weekly earnings of a full-time worker 25 years old or older for the years 1980 and 1997.



### Median Weekly Earnings

	Less than 4 Years of High School	High School Diploma	1 to 3 Years of College	College Degree
1980	\$222	\$266	\$304	\$376
1997	\$321	\$461	\$518	\$779

Source: *The Wall Street Journal 1999 Almanac*

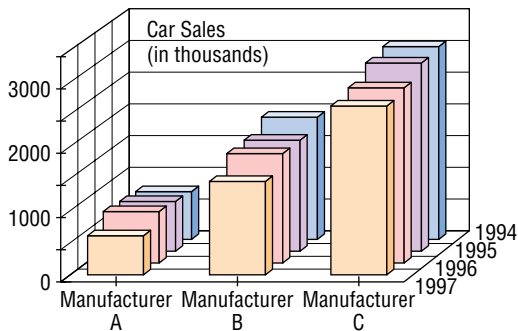
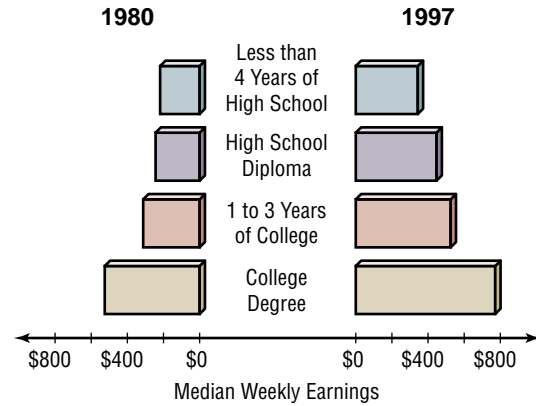
- Make a back-to-back bar graph that represents the data.
- Describe any trends indicated by the graph.

*(continued on the next page)*



a. Let the level of education be the central axis. Draw a horizontal axis that is scaled \$0 to \$800 in each direction. Let the left side of the graph represent the earnings from 1980 and the right side of the graph be those from 1997. Draw the bars to the appropriate length for the data.

b. You can see from the graph that when you compare each level of education with the next, more education resulted in a greater increase in median weekly earnings in 1997 than in 1980.



Sometimes it is desirable to show three aspects of a set of data at the same time. To present data in this way, a **three-dimensional bar graph** is often used. The graph at the left represents the retail sales in thousands of passenger cars in the United States for three major domestic car manufacturers during the years 1994 to 1997. The grid defines the car and year. The height of each bar represents the number of cars sold each year.

Sometimes the amount of data you wish to represent in a bar graph is too great for each item of data to be considered individually. In this case, a **frequency distribution** is a convenient system for organizing the data. A number of classes are determined, and all values in a class are tallied and grouped together. To determine the number of classes, first find the **range**. The range of a set of data is the difference between the greatest and the least values in the set.

Retail Management Testing Scores	
Scores	Frequency
60–70	9
70–80	10
80–90	12
90–100	3

The intervals are often named by a range of values. In the table, the interval described by 60–70 means all the test scores  $s$  such that  $60 \leq s < 70$ . The **class interval** is the range of each class. The class intervals in a frequency distribution should all be equal. In the table, the range for each class interval is 10.

The **class limits** of a set of data organized in a frequency distribution are the upper and lower values in each interval. The class limits in the testing data above are 60, 70, 80, 90, and 100. The **class marks** are the midpoints of the classes; that is the average of the upper and lower limit for each interval. The class mark for the interval 60–70 is  $\frac{60 + 70}{2}$  or 65.

The most common way of displaying frequency distributions is by using a **histogram**. A histogram is a type of bar graph in which the width of each bar represents a class interval and the height of the bar represents the frequency in that interval. Histograms usually have fewer than ten intervals.

*Class intervals are often multiples of 5.*

*The difference in consecutive class marks is the same as the class interval.*



**Example**



**2 FOOTBALL** The winning scores for the first 34 Super Bowls are 35, 33, 16, 23, 16, 24, 14, 24, 16, 21, 32, 27, 35, 31, 27, 26, 27, 38, 38, 46, 39, 42, 20, 55, 20, 37, 52, 30, 49, 27, 35, 31, 34, and 23.



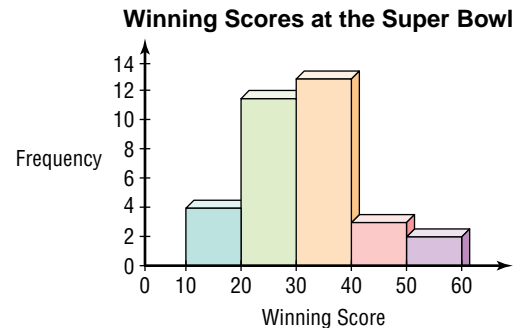
Vince Lombardi Trophy

- Find the range of the data.
- Determine an appropriate class interval.
- Find the class marks.
- Construct a frequency distribution of the data.
- Draw a histogram of the data.
- What conclusions can you determine from the graph?

- The range of the data is  $55 - 14$  or 41.
- An appropriate class interval is 10 points, beginning with 10 points and ending with 60 points. There will be five classes.
- The class marks are the averages of the class limits of each interval. The class marks are 15, 25, 35, 45, and 55.
- Make a table listing class limits. Use tallies to determine the number of scores in each interval.

Winning Score	Tallies	Frequency
10–20		4
20–30		12
30–40		13
40–50		3
50–60		2

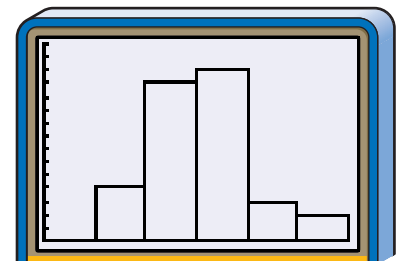
- Label the horizontal axis with the class limits. The vertical axis should be labeled from 0 to a value that will allow for the greatest frequency. Draw the bars side by side so that the height of each bar corresponds to its interval's frequency.



**Graphing Calculator Appendix**

For keystroke instruction on how to create a histogram, see pages A23-A24.

You can also use a graphing calculator to create the histogram. In statistics mode, enter the class marks in the **L1** list and the frequency in the **L2** list. Set the window using the class interval for **Xscl**, and select the histogram as the type of graph.

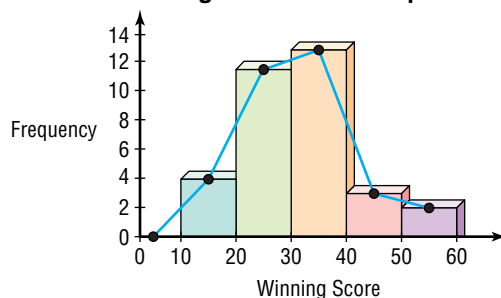


[0, 60] scl:10 by [0, 15] scl:1

- The winning score at the Super Bowl tends to be between 20 and 40 points.

Another type of graph can be created from a histogram. A broken line graph, often called a **frequency polygon**, can be drawn by connecting the class marks on the graph. The class marks are graphed as the midpoints of the top edge of each bar. The frequency polygon for the histogram in Example 2 is shown at the right.

Winning Scores at the Super Bowl



**Example**



**3 HEALTH** A graduate student researching the effect of smoking on blood pressure collected the following readings of systolic blood pressure from 30 people within a control group.

125, 145, 110, 126, 128, 180, 177, 176, 156, 144, 182, 205, 191, 140, 138, 126, 154, 163, 172, 159, 174, 151, 142, 160, 147, 143, 158, 129, 132, 137

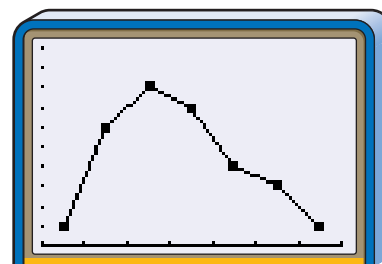
- Find an appropriate class interval. Then name the class limits and the class marks.
- Construct a frequency distribution.
- Use a graphing calculator to draw a frequency polygon.

a. The range of the data is  $205 - 110$  or 95. An appropriate class interval is 15 units. The class limits are 105, 120, 135, 150, 165, 180, 195, and 210. The class marks are 112.5, 127.5, 142.5, 157.5, 172.5, 187.5, and 202.5.

b.

Systolic Blood Pressure	Tallies	Frequency
105–120		1
120–135		6
135–150		8
150–165		7
165–180		4
180–195		3
195–210		1

- c. In statistics mode, enter the class marks in the L1 list and the frequency in the L2 list. Set the window using the class interval for Xscl, and select the line graph as the type of graph.



[105, 210] scl:15 by [0, 10] scl:1



**Graphing Calculator Appendix**

For keystroke instruction on how to create a frequency polygon, see page A24.

**CHECK FOR UNDERSTANDING**

**Communicating Mathematics**

Read and study the lesson to answer each question.

- Compare and contrast line plot, bar graph, histogram, and frequency polygon.
- Explain how to construct a frequency distribution.



3. **Determine** which class intervals would be appropriate for the data. Explain.  
55, 72, 51, 47, 73, 81, 74, 88, 83, 47, 58, 66, 64, 71, 73, 84, 61, 89, 73, 82  
a. 1                      b. 5                      c. 10                      d. 20                      e. 30
4. *Math Journal* **Select** three graphs from newspapers or magazines. For each graph, write what conclusions might be drawn from the graph.
5. **Population** The table gives the percent of the U.S. population by age group.

**Guided Practice**



**Percent of U.S. Population**

Age	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
<b>1900</b>	14.8%	14.1%	16.3%	16.8%	12.6%	8.8%	8.6%	8.0%
<b>1999</b>	14.2%	14.4%	13.3%	15.6%	15.2%	10.7%	7.3%	9.3%

Source: U.S. Bureau of the Census

- a. Make a back-to-back bar graph of the data.  
b. Describe any trends indicated by the graph.
6. **History** The ages of the first 42 presidents when they first took office are listed.  
57, 61, 57, 57, 58, 57, 61, 54, 68, 51, 49, 64, 50, 48, 65, 52, 56, 46, 54, 49, 50,  
47, 55, 55, 54, 42, 51, 56, 55, 51, 54, 51, 60, 62, 43, 55, 56, 61, 52, 69, 64, 46
- a. Find the range of the data.  
b. Determine an appropriate class interval.  
c. What are the class limits?  
d. Find the class marks.  
e. Construct a frequency distribution of the data.  
f. Draw a histogram of the data.  
g. Name the interval or intervals that describe the age of most presidents.

**EXERCISES**

**Applications and Problem Solving**



7. **Sales** As customers come to the cash register at an electronics store, the sales associate asks them to give their ZIP code. During one hour, a sales associate gets the following responses.  
43221, 43212, 43026, 43220, 43214, 43026, 43229, 43229, 43220, 45414,  
43220, 43221, 43212, 43220, 43212, 43220, 43221, 43221, 43214, 43026
- a. Make a line plot showing how many times each ZIP code was recorded.  
b. Which ZIP code was recorded most frequently?  
c. Why would a store want this type of information?
8. **Transportation** The average number of minutes men and women drivers spend behind the wheel daily is given below.

Age	16-19	20-34	35-49	50-64	65+
<b>Men</b>	58	81	86	88	73
<b>Women</b>	56	65	67	61	55

Source: Federal Highway Administration and the American Automobile Manufacturers Association

- a. Make a back-to-back bar graph of the data.  
b. What conclusions can you draw from the graph?



- 9. Entertainment** The table gives data on the rental revenue and the sale revenue of home videos as well as predictions of future revenues.

Year	Total Rental Revenue (in billions)	Total Sales Revenue (in billions)
1985	\$2.55	\$0.86
1990	\$6.63	\$3.18
1997	\$7.46	\$8.24
2000	\$9.18	\$9.76
2005	\$9.26	\$13.90

Source: Video Software Dealers Association

- Make a back-to-back bar graph of the data.
- Which market, rental or sales, seems to have a better future? Explain.

- 10. Nutrition** The grams of fat in various sandwiches served by national fast-food restaurants are listed below.

18, 27, 15, 23, 27, 14, 15, 19, 39, 53, 31, 29, 12, 43, 38, 4, 10, 9, 21,  
31, 31, 25, 28, 20, 22, 46, 15, 31, 16, 20, 30, 8, 18, 15, 7, 9, 5, 8

- What is the range of the data?
- Determine an appropriate class interval.
- Name the class limits.
- What are the class marks?
- Construct a frequency distribution of the data.
- Draw a histogram of the data.
- Name the interval or intervals that describe the fat content of most sandwiches.

- 11. Sports** The number of nations represented at the first eighteen Olympic Winter Games are listed below.

Year	Place	Number of Nations	Year	Place	Number of Nations
1924	Chamonix, France	16	1968	Grenoble, France	37
1928	St. Moritz, Switzerland	25	1972	Sapporo, Japan	35
1932	Lake Placid, U.S.A.	17	1976	Innsbruck, Austria	37
1936	Garmisch-Partenkirchen, Germany	28	1980	Lake Placid, U.S.A.	37
1948	St. Moritz, Switzerland	28	1984	Sarajevo, Yugoslavia	49
1952	Oslo, Norway	30	1988	Calgary, Canada	57
1956	Cortina D'Ampezzo, Italy	32	1992	Albertville, France	64
1960	Squaw Valley, U.S.A.	30	1994	Lillehammer, Norway	67
1964	Innsbruck, Austria	36	1998	Nagano, Japan	72

Source: *The Complete Book of the Olympics*

- Find the range of the data.
- What is an appropriate class interval?
- What are the class limits?
- What are the class marks?
- Construct a frequency distribution of the data.
- Draw a histogram of the data.
- Use the histogram to draw a frequency polygon.



12. **Architecture** The heights (in feet) of the tallest buildings in selected cities in the United States are listed below. Source: *The World Almanac*, 1999

1023 871 405 714 739 546 535 1000 626 1018  
738 697 500 440 945 450 471 943 579 404

- Find the range of the heights of these buildings.
  - What is an appropriate class interval?
  - Construct a frequency distribution of the data.
  - Draw a frequency polygon of the data.
  - Which interval or intervals represent the greatest number of these buildings?
13. **Baseball** The greatest numbers of stolen bases for a single player are listed.

**interNET**  
CONNECTION

**Data Update**  
For the latest information about stolen bases, visit [www.amc.glencoe.com](http://www.amc.glencoe.com)



**Greatest Number of Stolen Bases for a Single Player**

Year		'90	'91	'92	'93	'94	'95	'96	'97	'98	'99
Stolen Bases	American League	65	58	66	70	60	54	75	74	66	44
	National League	77	76	78	58	39	56	53	60	58	72

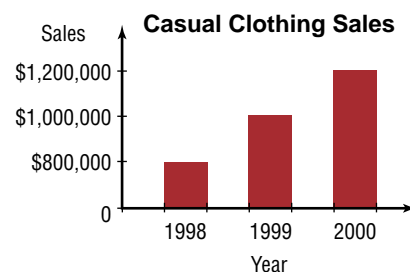
Source: *Information Please Almanac*, 1999

- Make a back-to-back bar graph for the data.
  - Combine data from both leagues to construct a frequency distribution.
  - Draw a histogram of the data. Then draw a frequency polygon.
  - How many players made 70 or more stolen bases to reach the record?
  - How many players made less than 50 stolen bases to reach the record?
14. **Critical Thinking** Create a set of data of 20 elements so that the data can be divided into five classes with class intervals of 0.5.
15. **Geography** The production of wheat, rice, and corn for 1997 is given in the table below. Make a three-dimensional bar graph of the data.

Country	Wheat (millions of tons)	Rice (millions of tons)	Corn (millions of tons)
China	122.6	198.5	105.4
India	68.7	123.0	9.8
United States	68.8	8.1	237.9

Source: UN Food and Agriculture Organization

16. **Critical Thinking** The graph shows a store's annual sales.
- Why is the graph misleading?
  - Draw a graph of the data that is not misleading.
  - Find a graph in a newspaper or magazine that you believe is misleading. Explain why you believe it is misleading.





17. **Statistics** Design a survey and ask your classmates to respond to the survey. Make an appropriate graph to depict the results of the survey.

**Mixed Review**

18. **Horticulture** The survival rate of a variety of mums in a certain area of the country is 80%. If 8 mums are planted, what is the probability that exactly 6 will survive? (*Lesson 13-6*)
19. Find the second term of  $(c - 2d)^7$ . (*Lesson 12-6*)
20. Solve  $3.6^x = 58.9$  by using logarithms. (*Lesson 11-6*)
21. Graph  $9xy = -36$ . (*Lesson 10-4*)
22. **SAT Practice Grid-In** If  $x^2 + y^2 = 16$  and  $xy = 8$ , what is  $(x + y)^2$ ?

## CAREER CHOICES

### Accountant



Everyone seems to like money. If you choose a career in accounting, you will be working with money, but only on paper. Accountants work with the financial records of individuals, businesses, or governments and prepare statements showing income and expenses. Accountants also prepare reports, including tax reports.

As an accountant, you may choose to specialize in an area that might include auditing or systems and procedures, or you may specialize in a particular business such as agriculture. To become a certified public accountant, you must pass an examination to be certified by the state. Accountants can work for businesses, for government, or work independently for any individual or business that desires their services.

#### CAREER OVERVIEW

**Degree Preferred:**

Bachelor's degree in accounting

**Related Courses:**

mathematics, communications, computer science, business courses

**Outlook:**

number of jobs expected to increase faster than the average through 2006



#### Consolidated Balance Sheets

	July 4 1999	January 3 1999
<b>Assets</b> (in thousands)		
Cash and cash equivalents	\$ 245,159	\$ 160,743
Accounts and notes receivable, net	78,865	94,689
Deferred income taxes	19,592	23,177
Inventories and other	35,181	35,085
<b>Total current assets</b>	<b>378,797</b>	<b>313,694</b>



For more information on careers in accounting, visit [www.amc.glencoe.com](http://www.amc.glencoe.com)



# Measures of Central Tendency

## OBJECTIVES

- Find the mean, median, and mode of a set of data.
- Find measures of central tendency of data organized in a stem-and-leaf plot or a frequency distribution table.



**ENTERTAINMENT** For many years, people have attended Broadway plays in New York City.

Broadway became an important theatrical district in the mid-1800s. The theatrical activity of Broadway peaked in the 1920s. Today, there are fewer new shows, but Broadway still remains a major theatrical center. The numbers of new Broadway productions in recent seasons are listed below. What is the average number of new Broadway productions for these seasons?

*This problem will be solved in Example 3.*



Season	New Productions	Season	New Productions	Season	New Productions
1960–1961	48	1973–1974	43	1986–1987	41
1961–1962	53	1974–1975	54	1987–1988	32
1962–1963	54	1975–1976	55	1988–1989	30
1963–1964	63	1976–1977	54	1989–1990	35
1964–1965	67	1977–1978	42	1990–1991	28
1965–1966	68	1978–1979	50	1991–1992	37
1966–1967	69	1979–1980	61	1992–1993	33
1967–1968	74	1980–1981	60	1993–1994	37
1968–1969	67	1981–1982	48	1994–1995	29
1969–1970	62	1982–1983	50	1995–1996	38
1970–1971	49	1983–1984	36	1996–1997	37
1971–1972	55	1984–1985	33	1997–1998	33
1972–1973	55	1985–1986	33	1998–1999	20

Source: The League of American Theatres and Producers, Inc.

The average number of new Broadway productions is an ambiguous term. Loosely stated, the average means the center of the distribution or the most typical case. Measures of average are also called **measures of central tendency** and include the **mean, median, and mode**.

The **arithmetic mean**  $\bar{X}$  is often referred to as the mean. The mean is found by adding the values in a set of data and dividing the sum by the number of values in that set. Every number in a set of data affects the value of the mean. Consequently, the mean is generally a good representative measure of central tendency. However, the mean can be considerably influenced by extreme values.



**Example 1** Find the mean of the set {19, 21, 18, 17, 18, 22, 46}.

$$\bar{X} = \frac{\text{sum of the values in the set of data}}{\text{number of values in the set}}$$

$$\bar{X} = \frac{19 + 21 + 18 + 17 + 18 + 22 + 46}{7}$$

$$\bar{X} = \frac{161}{7} \text{ or } 23$$

The mean of the set of data is 23.

*Notice that the mean is not necessarily a member of the set of data.*

The general formula for the mean of any set of data can be written using sigma notation. If  $X$  is a variable used to represent any value in a set of data containing  $n$  items, then the arithmetic mean  $\bar{X}$  of  $n$  values is given by the following formula.

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

The numerator of the fraction can be abbreviated using the summation symbol  $\Sigma$ . *Recall that  $\Sigma$  is the uppercase Greek letter sigma.*

$$\sum_{i=1}^n X_i = X_1 + X_2 + X_3 + \dots + X_n$$

The symbol  $X_i$  represents successive values of the set of data as  $i$  assumes successive integral values from 1 to  $n$ . Substitute the sigma notation into the formula for the mean to obtain the formula below.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \text{ or } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

### Arithmetic Mean

If a set of data has  $n$  values given by  $X_i$  such that  $i$  is an integer and  $1 \leq i \leq n$ , then the arithmetic mean  $\bar{X}$  can be found as follows.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Another measure of central tendency is the median, symbolized by  $M_d$ .

### Median

The median of a set of data is the middle value. If there are two middle values, it is the mean of the two middle values.

*Notice that the median is not necessarily a member of the set of data.*

Before the median can be found, the data must be arranged in an ordered sequence, usually from least to greatest. The median of the set {5, 6, 8, 11, 14} is the middle value 8. The median of the set {3, 4, 6, 7, 8, 10} is  $\frac{6+7}{2}$  or 6.5.

The median is preferable to the mean as a measure of central tendency when there are a few extreme values or when some of the values cannot be determined. Unlike the mean, the median is influenced very little by extreme values.



The mode of a set of data is the most frequent value. Some sets of data have multiple modes and others have no mode.

Data with two modes are **bimodal**. Sets have no mode when each item of the set has equal frequency. The value of the mode is not affected by extreme values. Unlike the mean and median, the mode, if it exists, is always a member of the set of data.

**Example**



**2 IMMIGRATION** The numbers of immigrants entering the United States in a recent year are given at the right.

- Find the mean of the data.
- Find the median of the data.
- Find the mode of the data.
- State which measures of central tendency seem most representative of the set of data. Explain.



**Immigrants Entering the United States**

Country	Number of Immigrants
China	41,700
Cuba	26,500
Dominican Republic	39,600
India	44,900
Jamaica	19,100
Mexico	163,600
Philippines	55,900
Russia	19,700

Source: U.S. Immigration and Naturalization Service

- Since there are 8 countries,  $n = 8$ .

$$\frac{1}{8} \sum_{i=1}^8 X_i = \frac{1}{8} (41,700 + 26,500 + 39,600 + 44,900 + 19,100 + 163,600 + 55,900 + 19,700) \text{ or } 51,375$$

The mean is 51,375.

- To find the median, order the data. Since all the numbers are multiples of 100, you can order the set by hundreds.

191      197      265      396      417      449      559      1636

Since there are an even number of data, the median is the mean of the two middle numbers, 39,600 and 41,700. The median number of immigrants is  $\frac{39,600 + 41,700}{2}$  or 40,650.

- Since all elements in the set of data have the same frequency, there is no mode.
- Notice that the mean is affected by the extreme value 163,600 and does not accurately represent the data. The median is a more representative measure of central tendency in this case.

When you have a large number of data, it is often helpful to use a **stem-and-leaf plot** to organize your data. In a stem-and-leaf plot, each item of data is separated into two parts that are used to form a **stem** and a **leaf**. The parts are organized into two columns.

**Stems:** The column on the left shows the stems. Stems usually consist of the digits in the greatest common place value of all the data. For example, if the set of data includes the numbers 890 and 1160, the greatest common place value is hundreds. Therefore, the stem of 890 is 8, and the stem of 1160 is 11.

**Leaves:** The column at the right contains the leaves. The leaves are one-digit numbers, which are in the next greatest place value after the stem. The leaf of 890 is 9, and the leaf of 1160 is 6. The stems and leaves are usually arranged from least to greatest.

**Example**

**3**

**ENTERTAINMENT** Refer to the application at the beginning of the lesson.



- Make a stem-and-leaf plot of the number of new productions for the seasons listed.
- Find the mean of the data.
- Find the median of the data.
- Find the mode of the data.
- What is a good representative number for the average of the new Broadway productions for the seasons 1960-1999?



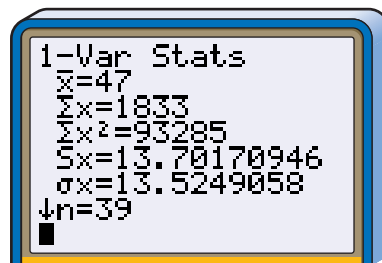
- Since the number of productions range from 20 to 74, we will use the tens place for the stems. List the stems and draw a vertical line to the right of the stems. Then list the leaves, which in this case will be the ones digit. As shown below, it is often helpful to list the leaves as you come to them and then rewrite the plot with the leaves in order from the least to greatest.

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

$$2|8 = 28$$

*An annotation usually accompanies a stem-and-leaf plot to give meaning to the representation.*

- Enter the data in the L1 list of a graphing calculator. Use the statistics mode of the calculator to find  $\bar{X}$ . The mean is 47.
- Since the median is the middle value, it is the 20th leaf on the plot. The median is 48.
- The stem-and-leaf plot shows the modes by repeated digits for a particular stem. There are four 3s with the stem 3. The mode is 33.



- e. Although 33 is the most common number in the data, it is not a central number for the data. In this case, the mean and median seem to be more representative of the data. Therefore, a representative average number of new Broadway productions could be either 47 or 48.

In a frequency distribution containing large amounts of data, each individual value in the set of data loses its identity. The data in each class are assumed to be uniformly distributed over the class. Thus, the class mark is assumed to be the mean of the data tallied in its class. For example, the mean of the data in the class with limits 17.5-22.5 is assumed to be 20, the class mark.

In the frequency distribution, the sum of the values in a class is found by multiplying the class mark  $X$  by the frequency  $f$  of that class. The sum of all the values in a given set of data is found by adding the sums of the values of each class in the frequency distribution. The sum of all values in the set can be

*Remember that the measures of central tendency are only representations of the set of data.*

represented by  $\sum_{i=1}^k (f_i \cdot X_i)$ , where  $k$  is the number of classes in the frequency distribution. Thus, the arithmetic mean of  $n$  values in a frequency distribution is found by dividing the sum of the values in the set by  $n$  or an expression equivalent to  $n$ , such as  $\sum_{i=1}^k f_i$ .

### Mean of the Data in a Frequency Distribution

If  $X_1, X_2, \dots, X_k$  are the class marks in a frequency distribution with  $k$  classes and  $f_1, f_2, \dots, f_k$  are the corresponding frequencies, then the arithmetic mean  $\bar{X}$  can be approximated as follows.

$$\bar{X} = \frac{\sum_{i=1}^k [f_i \cdot X_i]}{\sum_{i=1}^k f_i}$$

### Example



- 4 EDUCATION** Estimate the mean of the scores of 100 students on an algebra test given the following frequency distribution.

Class Limits	Class Marks ( $X$ )	Frequency ( $f$ )	$fX$
97.5-102.5	100	5	500
92.5-97.5	95	9	855
87.5-92.5	90	17	1530
82.5-87.5	85	26	2210
77.5-82.5	80	22	1760
72.5-77.5	75	10	750
67.5-72.5	70	7	490
62.5-67.5	65	4	260

$$\sum_{i=1}^8 f_i = 100 \quad \sum_{i=1}^8 (f_i \cdot X_i) = 8355$$

$$\bar{X} = \frac{8355}{100} \text{ or } 83.55$$

The mean is approximately 84.



The median  $M_d$  of the data in a frequency distribution is found from the **cumulative frequency distribution**. The cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the previous classes. The chart shows the cumulative frequency for the data in Example 4.

*It is often helpful to calculate the cumulative frequency from the last interval to the first.*

Class Limits	Frequency $f$	Cumulative Frequency
97.5–102.5	5	100
92.5–97.5	9	95
87.5–92.5	17	86
82.5–87.5	26	69
77.5–82.5	22	43
72.5–77.5	10	21
67.5–72.5	7	11
62.5–67.5	4	4

For the class limit 77.5–82.5, the cumulative frequency equals  $22 + 10 + 7 + 4$  or 43. This means that 43 algebra test scores fall below 82.5.

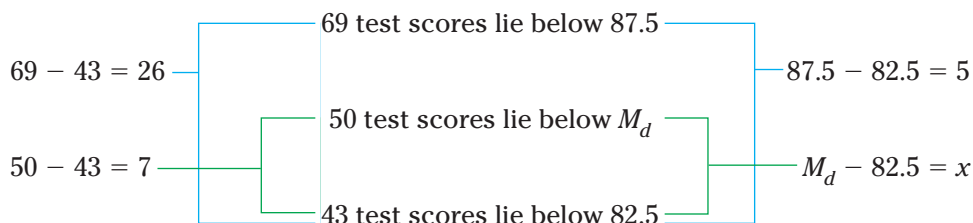
Since the median is the value below which 50% of the data lie, the class in which the median lies can be located. This class is called the **median class**. The median can be found by using an estimation technique called interpolation. *This method can also be used to find a score at any percent level.*

**Example 5** **EDUCATION** Estimate the median of the data in the frequency distribution in Example 4.



Since there are 100 scores in this frequency distribution, 50 scores are below the median and 50 are above. From the chart above, find the least cumulative frequency that is greater than or equal to 50. That cumulative frequency is 69. So, the median class is 82.5–87.5.

You can use a proportion to find the value of  $M_d$  by finding the ratios of the differences in the cumulative frequencies and the upper limits of the classes.



$$\frac{26}{5} = \frac{7}{x}$$

$$x \approx 1.346153846 \quad \text{Use a calculator.}$$

$$M_d - 82.5 = x$$

$$M_d - 82.5 \approx 1.3 \quad x \approx 1.3$$

$$M_d \approx 83.8$$

The median of the data is approximately 83.8.



## CHECK FOR UNDERSTANDING

### Communicating Mathematics

Read and study the lesson to answer each question.

1. **Name** three measures of central tendency. Explain how to determine each measure.
2. **Write** a set of data that is bimodal.
3. **Explain** how to make a stem-and-leaf plot for a set of data whose greatest value is 1430 and least value is 970.
4. **You Decide** Eight people volunteer at Central City Recreation Center. The numbers of hours the volunteers spent at the center last week are 2, 3, 15, 2, 2, 23, 19, and 2. Omar says the median is the best representative for the number of volunteer hours. Tia disagrees and claims that the mean is a better representation. Casey says the most representative number is the mode. Who is correct? Explain.

### Guided Practice

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

5. {10, 45, 58, 10}
6. {24, 28, 21, 37, 31, 29, 23, 22, 34, 31}

7. Find the mean, median, and mode of the data represented by the stem-and-leaf plot at the right.

stem	leaf
9	1 4 5 8 9
10	5 5 7 7 7
11	1 1 2
<i>9   1 = 9100</i>	

8. The Laketown Senior Center recorded how many times its members use the center's fitness facilities during a typical month. They organized their data into the following chart.

Visits	1-5	5-9	9-13	13-17	17-21	21-25	25-29	29-33
Members	2	8	15	6	38	31	13	7

- a. How many members used the center's fitness facilities during the month?
  - b. Estimate the mean of the data.
  - c. What is the median class of the data?
  - d. Estimate the median of the data.
9. **Football** Each December, the Liberty Bowl is played in Memphis, Tennessee. The winning scores of the first 40 Liberty Bowl games are listed below.

7, 41, 15, 6, 16, 32, 13, 14, 14, 34, 47, 17, 7, 31, 31, 7, 20, 36, 21, 20,  
9, 28, 31, 21, 19, 21, 21, 21, 20, 34, 42, 23, 38, 13, 18, 30, 19, 41, 41, 23

- a. Make a stem-and-leaf plot of the winning scores.
- b. What is the mean of the data?
- c. What is the median of the data?
- d. Find the mode of the data.
- e. What is the most representative measure of central tendency for the number of points scored by the winning team at the Liberty Bowl? Explain.





## EXERCISES

### Practice

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

10. {140, 150, 160, 170}                      11. {3, 3, 6, 12, 3}
12. {21, 19, 17, 19}                          13. {5, 8, 18, 5, 3, 18, 14, 15}
14. {64, 87, 62, 87, 63, 98, 76, 54, 87, 58, 70, 76}
15. {6, 9, 11, 11, 12, 7, 6, 11, 5, 8, 10, 6}

16. Crates of books are being stored for later use. The weights of the crates in pounds are 142, 160, 151, 139, 145, 117, 172, 155, and 124.

- a. What is the mean of their weights?
- b. Find the median of their weights.
- c. If 5 pounds is added to each crate, how will the mean and median be affected?

Find the mean, median, and mode of the data represented by each stem-and-leaf plot.

17. 

stem	leaf
3	5 8 8 9
4	4 5 5 5 8
5	7 7 9
3	5 = 35

18. 

stem	leaf
5	2 4 6
6	0 1 7 8 9
7	1 6
8	0 2 6
9	1
5	2 = 5.2

19. 

stem	leaf
9	0 1 7 8 9
10	5 6 9
11	3 8 8 8
12	0 5 5
9	0 = 900

20. Make a stem-and-leaf plot of the following ages of people attending a family picnic.

- 15, 55, 35, 46, 28, 35, 25, 17, 30, 30, 27, 35,  
15, 25, 25, 20, 20, 15, 20, 17, 15, 25, 10

21. The store manager of a discount department store is studying the weekly wages of the part-time employees. The table profiles the employees.

- a. Find the sum of the wages in each class.
- b. What is the sum of all of the wages in the frequency distribution?
- c. Find the number of employees in the frequency distribution.
- d. What is the mean weekly wage in the frequency distribution?
- e. Find the median class of the frequency distribution.
- f. Estimate the median weekly wage in the frequency distribution.
- g. Explain why both the mean and median are good measures of central tendency in this situation.

Weekly Wages	Frequency
\$130–\$140	11
\$140–\$150	24
\$150–\$160	30
\$160–\$170	10
\$170–\$180	13
\$180–\$190	8
\$190–\$200	4

22. Find the value of  $x$  so that the mean of {2, 4, 5, 8,  $x$ } is 7.5.
23. What is the value of  $x$  so that the mean of { $x$ ,  $2x - 1$ ,  $2x$ ,  $3x + 1$ } is 6?
24. Find the value of  $x$  so that the median of {11, 2, 3, 3.2, 13, 14, 8,  $x$ } is 8.



25. The frequency distribution of the verbal scores on the SAT test for students at Kennedy High School is shown below.

Scores	Number of Students	Scores	Number of Students
200–250	9	500–550	18
250–300	14	550–600	12
300–350	23	600–650	7
350–400	30	650–700	3
400–450	33	700–750	1
450–500	28	750–800	1

- What is the mean of the verbal scores at Kennedy High School?
- What is the median class of the frequency distribution?
- Estimate the median of the verbal scores at Kennedy High School.

**Applications  
and Problem  
Solving**



26. **Weather** The growing season in Tennessee is the period from May to September. The table at the right shows the normal rainfall for those months.

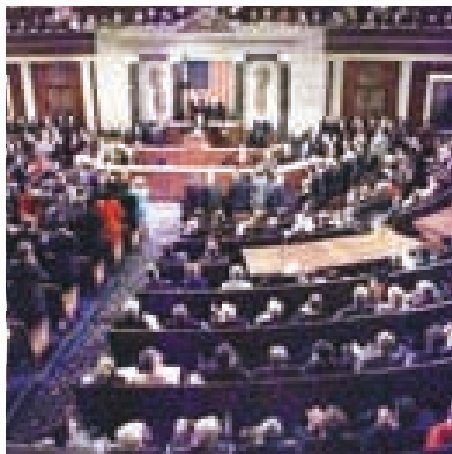
Normal Rainfall for Tennessee (inches)	
May	4.8
June	3.6
July	3.9
August	3.6
September	3.7

- Find the mean, median, and mode of this data.
- Suppose Tennessee received heavy rain in May totaling 8.2 inches. If this figure were used for May, how would the measures of central tendency be affected?
- If September were eliminated from the period, how would this affect the measures of central tendency?

27. **Critical Thinking** Find a set of numbers that satisfies each list of conditions.

- The mean, median, and mode are all the same number.
- The mean is greater than the median.
- The mode is 10 and the median is greater than the mean.
- The mean is 6, the median is  $5\frac{1}{2}$ , and the mode is 9.

28. **Government** As of 1999, the number of members in the House of Representatives for each state is given below.



AL	7	HI	2	MA	10	NM	3	SD	1
AK	1	ID	2	MI	16	NY	31	TN	9
AZ	6	IL	20	MN	8	NC	12	TX	30
AR	4	IN	10	MS	5	ND	1	UT	3
CA	52	IA	5	MO	9	OH	19	VT	1
CO	6	KS	4	MT	1	OK	6	VA	11
CT	6	KY	6	NE	3	OR	5	WA	9
DE	1	LA	7	NV	2	PA	21	WV	3
FL	23	ME	2	NH	2	RI	2	WI	9
GA	11	MD	8	NJ	13	SC	6	WY	1

- Make a stem-and-leaf plot of the number of representatives.
- Find the mean of the data.
- What is the median of the data?
- Find the mode of the data.
- What is a representative average for the number of members in the House of Representatives per state? Explain.

**Data Update**

For the latest information about the number of goals scored in hockey, visit [www.amc.glencoe.com](http://www.amc.glencoe.com)



**29. Hockey** A frequency distribution for the number of goals scored by teams in the National Hockey League during a recent season are given at the right.

- Use the frequency chart to estimate the mean of the number of goals scored by a team.
- What is the median class of the frequency distribution?
- Use the frequency chart to estimate the median of the number of goals scored by a team.



**National Hockey League Goals**

Goals	Number of Teams
160–180	1
180–200	6
200–220	10
220–240	6
240–260	3
260–280	1

Source: National Hockey League

- The actual numbers of goals scored are listed below. Find the mean and median of the data.

268, 248, 245, 242, 239, 239, 237, 236, 231, 230, 217, 215, 214, 211, 210, 210, 207, 205, 202, 200, 196, 194, 192, 190, 189, 184, 179

- How do the measures of central tendency found by using the frequency chart compare with the measures of central tendency found by using the actual data?

**30. Critical Thinking** A one-meter rod is suspended at its middle so that it balances. Suppose one-gram weights are hung on the rod at the following distances from one end.

5 cm    20 cm    37 cm    44 cm    52 cm    68 cm    71 cm    85 cm

The rod does not balance at the 50-centimeter mark.

- Where must a one-gram weight be hung so that the rod will balance at the 50-centimeter mark?
- Where must a two-gram weight be hung so that the rod will balance at the 50-centimeter mark?

**31. Salaries** The salaries of the ten employees at the XYZ Corporation are listed below.

\$54,000, \$75,000, \$55,000, \$62,000, \$226,000,  
\$65,000, \$59,000, \$61,000, \$162,000, \$59,000

- What is the mean of the salaries?
- Find the median of the salaries.
- Find the mode of the salaries.
- What measure of central tendency might an employee use when asking for a raise?
- What measure of central tendency might management use to argue against a raise for an employee?
- What measure of central tendency do you think is most representative of the data? Why?
- Suppose you are an employee of the company making \$75,000. Write a convincing argument that you deserve a raise.

- 32. Education** The grade point averages for a graduating class are listed in the frequency table below.

Grade Point Averages	1.75–2.25	2.25–2.75	2.75–3.25	3.25–3.75	3.75–4.25
Frequency	12	15	31	37	5

- a. What is the estimated mean of the data?  
 b. Estimate the median of the data.
- 33. Basketball** Jackson High School just announced the members of its varsity basketball team for the year. Kwan, who is 5' 9" tall, is the only sophomore to make the team. The other basketball team members are 5' 11", 6' 0", 5' 7", 6' 3", 6' 1", 6' 6", 5' 8", 5' 9" and 6' 2". How does Kwan compare with the other team members?

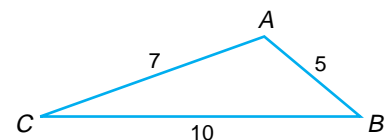
**Mixed Review**

- 34. Highway Safety** The maximum speed limits in miles per hour for interstate highways for the fifty states are given below. Construct a frequency polygon of the data. (Lesson 14-1)
- 70, 65, 75, 70, 70, 75, 65, 65, 70, 70, 55, 75, 65, 65, 65, 70, 65, 70, 65, 65, 65, 70, 70, 70, 70, 65, 75, 75, 65, 65, 75, 65, 70, 70, 65, 75, 65, 65, 65, 65, 75, 65, 70, 75, 65, 65, 70, 70, 65, 75



Source: National Motorists Association

- 35.** Determine if the following event is *independent* or *dependent*. Then determine the probability. (Lesson 13-4)
- the probability of randomly selecting two fitness magazines at one time from a basket containing 6 news magazines, 3 fitness magazines, and 2 sports magazines
- 36.** Use the ratio test to determine if the series  $\frac{1}{3} + \frac{2}{3^2} + \frac{3}{3^3} + \dots + \frac{n}{3^n} + \dots$  is *convergent* or *divergent*. (Lesson 12-4)
- 37. Investments** An annuity pays 6%. What is the future value of the annuity if \$1500 is deposited into the account every 6 months for 10 years? (Lesson 11-2)
- 38.** Graph the system of inequalities. (Lesson 10-8)
- $$3x + y^2 \leq 18$$
- $$x^2 + y^2 \geq 9$$
- 39. SAT Practice** The area of  $\triangle ABC$  is between which pair of numbers?
- A 16 and 17  
 B 15 and 16  
 C 12 and 13  
 D 10 and 11  
 E 9 and 10



# Measures of Variability

## OBJECTIVES

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



**EDUCATION** Are you planning to attend college? If so, do you know which school you are going to attend? There are several factors influencing students' decisions concerning which college to attend.

Two of those factors may be the cost of tuition and the size of the school. The table lists some of the largest colleges with their total enrollment and cost for in-state tuition and fees.



## College Enrollment and Tuition

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

Source: College Entrance Examination Board

## interNET CONNECTION

### Data Update

For the latest information about college enrollment and tuition, visit [www.amc.glencoe.com](http://www.amc.glencoe.com)

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

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

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

$$47,476 - 17,090 = 30,386.$$

University of Texas

University of Alaska

The range of enrollment is 30,386 students.



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

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

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

### Semi-Interquartile Range

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

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

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



#### Graphing Calculator Tip

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

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

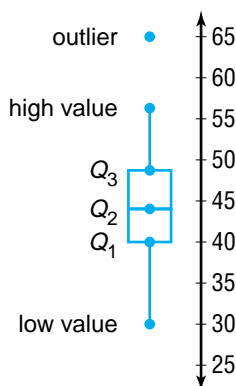
b. Find the semi-interquartile range of the college enrollments.

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

17,090 18,001 19,886 20,325  $\underline{20,925}$   $Q_1$  21,498 21,645 21,970 22,393  $Q_2$   $\underline{22,836}$   
 $Q_3$   
 24,075 24,141 24,481 27,874  $\underline{28,285}$  29,693 37,718 45,462 47,476

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

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



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

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

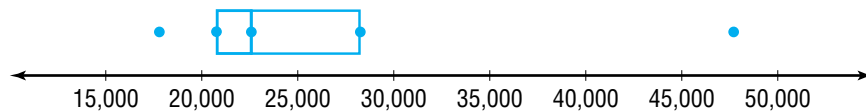
**Example**



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

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

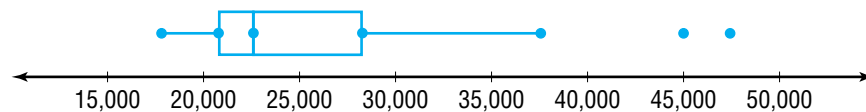
Draw a number line and plot the quartiles, the median, and the extreme values. Draw a box to show the interquartile range. Draw a segment through the median to divide the box into two smaller boxes.



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

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

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



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

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

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

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



## Mean Deviation

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

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

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

### Example



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

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

#### Method 1: Sigma notation

$$\begin{aligned} MD &\approx \frac{1}{19} \sum_{i=1}^{19} |X_i - 26,093.37| \\ MD &\approx \frac{1}{19} (|47,476 - 26,093.37| + |45,462 - 26,093.37| + \dots + |17,090 - 26,093.37|) \\ MD &\approx \frac{1}{19} (|21,382.63| + |19,368.63| + \dots + |-9003.37|) \\ MD &\approx 6310.29 \end{aligned}$$



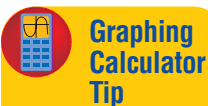
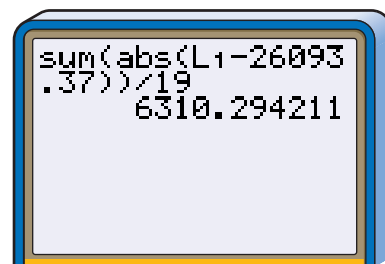
The mean deviation of the enrollments is about 6310.29. This means that the enrollments are an average of about 6310.29 above or below the mean enrollment of 26,093.37.

#### Method 2: Graphing Calculator

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

$$\text{sum}(\text{abs}(L1 - 26093.37))/19$$

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



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

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



## Standard Deviation

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

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$\sigma$  is the lowercase Greek letter sigma.

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

### Example



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

#### Method 1: Standard Deviation Formula

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

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

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

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

$$\sigma \approx 8354.59$$

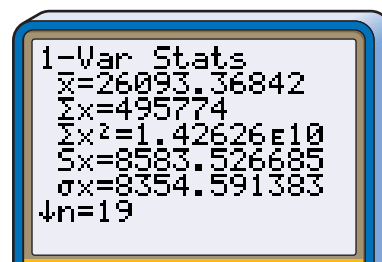
The standard deviation is about 8354.59. Since the mean of the enrollments is about 26,093.37 and the standard deviation is about 8354.59, the data have a great amount of variability.

#### Method 2: Graphing Calculator

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

The standard deviation, indicated by  $\sigma x$ , is the fifth statistic listed.

The mean ( $\bar{x}$ ) is 26,093.36842 and the standard deviation is 8354.591383, which agree with the calculations using the formulas.



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

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

**Standard  
Deviation of  
the Data in a  
Frequency  
Distribution**

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

$$\sqrt{\frac{\sum_{i=1}^k [X_i - \bar{X}]^2 \cdot f_i}{\sum_{i=1}^k f_i}}$$

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

**Example**



**5**

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

**Method 1: Using Formulas**

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

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

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

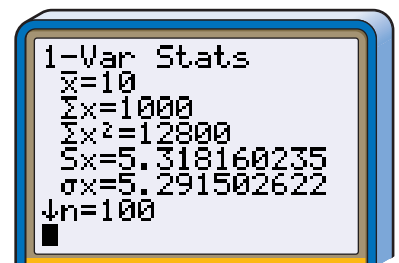
Since the mean number of price-earnings ratios is 10 and the standard deviation is 5.29, this indicates a great amount of variability in the data.

**Method 2: Graphing Calculator**

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

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

The calculator confirms the standard deviation is about 5.29.

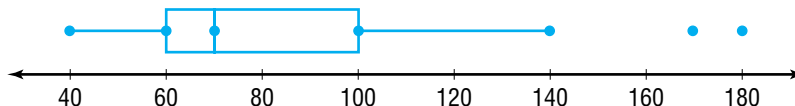


## CHECK FOR UNDERSTANDING

### Communicating Mathematics

Read and study the lesson to answer each question.

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



2. **Explain** how to find the variance of a set of data if you know the standard deviation.
3. **Compare and contrast** mean deviation and standard deviation.
4. *Math Journal* **Draw** a box-and-whisker plot for data you found in a newspaper or magazine. What conclusions can you derive from the plot?
5. Find the interquartile range and the semi-interquartile range of {17, 28, 44, 37, 28, 42, 21, 41, 35, 25}. Then draw a box-and-whisker plot.
6. Find the mean deviation and the standard deviation of {\$4.45, \$5.50, \$5.50, \$6.30, \$7.80, \$11.00, \$12.20, \$17.20}
7. Find the arithmetic mean and the standard deviation of the frequency distribution at the right.

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

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

Normal Maximum Daily Temperatures						
	January	February	March	April	May	June
Los Angeles	65.7	65.9	65.5	67.4	69.0	71.9
Las Vegas	57.3	63.3	68.8	77.5	87.8	100.3
	July	August	September	October	November	December
Los Angeles	75.3	76.6	76.6	74.4	70.3	65.9
Las Vegas	105.9	103.2	94.7	82.1	67.4	57.5

Source: National Oceanic and Atmosphere Administration

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

# EXERCISES

## Practice

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

9. {30, 28, 24, 24, 22, 22, 21, 17, 16, 15}
10. {7, 14, 18, 72, 13, 15, 19, 8, 17, 28, 11, 15, 24}
11. {15.1, 9.0, 8.5, 5.8, 6.2, 8.5, 10.5, 11.5, 8.8, 7.6}
12. Use a graphing calculator to draw a box-and-whisker plot for {7, 1, 11, 5, 4, 8, 12, 15, 9, 6, 5, 9}?



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

13. {200, 476, 721, 579, 152, 158}
14. {5.7, 5.7, 5.6, 5.5, 5.3, 4.9, 4.4, 4.0, 4.0, 3.8}
15. {369, 398, 381, 392, 406, 413, 376, 454, 420, 385, 402, 446}
16. Find the variance of {34, 55, 91, 13, 22}.

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

17.

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

18.

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

19.

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

## Applications and Problem Solving



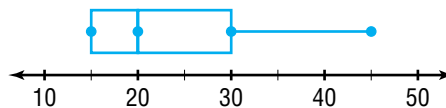
20. **Geography** There are seven navigable rivers that feed into the Ohio River. The lengths of these rivers are given at the right.

- a. Find the median of the lengths.
- b. Name the first quartile point and the third quartile point.
- c. Find the interquartile range.
- d. What is the semi-interquartile range?
- e. Are there any outliers? If so, name them.
- f. Make a box-and-whisker plot of the lengths of the rivers.
- g. Use the box-and-whisker plot to discuss the variability of the data.

Length of Rivers Feeding into the Ohio River	
Monongahela	129 miles
Allegheny	325 miles
Kanawha	97 miles
Kentucky	259 miles
Green	360 miles
Cumberland	694 miles
Tennessee	169 miles

Source: *The Universal Almanac*

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



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



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

Source: The National Collegiate Athletic Association

- What is the median of the number of women's teams playing a sport?
  - Find the first quartile point and the third quartile point.
  - What is the interquartile range and semi-interquartile range?
  - Are there any outliers? If so, name them.
  - Make a box-and-whisker of the number of women's teams playing a sport.
  - What is the mean of the number of women's teams playing a sport?
  - Find the mean deviation of the data.
  - Find the variance of the data.
  - What is the standard deviation of the data?
  - Discuss the variability of the data.
23. **Education** Refer to the data on the college tuition and fees in the application at the beginning of the lesson.
- What are the quartile points of the data?
  - Find the interquartile range.
  - Name any outliers.
  - Make a box-and-whisker plot of the data.
  - What is the mean deviation of the data?
  - Find the standard deviation of the data.
  - Discuss the variability of the data.
24. **Government** The number of times the first 42 presidents vetoed bills are listed below.
- 2, 0, 0, 7, 1, 0, 12, 1, 0, 10, 3, 0, 0, 9, 7, 6, 29, 93, 13, 0, 12, 414, 44, 170, 42, 82, 39, 44, 6, 50, 37, 635, 250, 181, 21, 30, 43, 66, 31, 78, 44, 25
- Make a box-and-whisker plot of the number of vetoes.
  - Find the mean deviation of the data.
  - What is the variance of the data?
  - What is the standard deviation of the data?
  - Describe the variability of the data.
25. **Entertainment** The frequency distribution shows the average audience rating for the top fifty network television shows for one season.

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

Source: Nielsen Media Research

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



26. **Critical Thinking** Is it possible for the variance to be less than the standard deviation for a set of data? If so, explain when this will occur. When would the variance be equal to the standard deviation for a set of data?
27. **Research** Find the number of students attending each school in your county. Make a box-and-whisker plot of the data. Determine various measures of variability and discuss the variability of the data.

**Mixed Review**

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

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

$5 | 4 = 5.4$

- a. What is the mean of the data?  
 b. Find the median of the data.  
 c. What is the mode of the data?

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

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

- a. Find the range of the number of programs sold.  
 b. Determine an appropriate class interval.  
 c. What are the class limits?  
 d. Construct a frequency distribution of the data.  
 e. Draw a histogram of the data.



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

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

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

- A 2 ft  $6\frac{1}{3}$  in.      B 2 ft  $8\frac{1}{3}$  in.      C 2 ft 7 in.  
 D 2 ft 8 in.      E 2 ft 9 in.

**MID-CHAPTER QUIZ**

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

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

- What is an appropriate class interval for the test scores? (Lesson 14-1)
- Construct a frequency distribution of the test scores. (Lesson 14-1)
- Draw a histogram of the test scores. (Lesson 14-1)
- Make a stem-and-leaf plot of the test scores. (Lesson 14-2)
- What is the mean of the test scores? (Lesson 14-2)
- Find the median of the test scores. (Lesson 14-2)
- Find the mode of the test scores. (Lesson 14-2)
- Make a box-and-whisker plot of the test scores. (Lesson 14-3)
- What is the mean deviation of the test scores? (Lesson 14-3)
- Discuss the variability of the data. (Lesson 14-3)

# The Normal Distribution

## OBJECTIVES

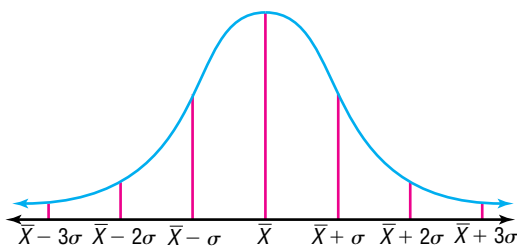
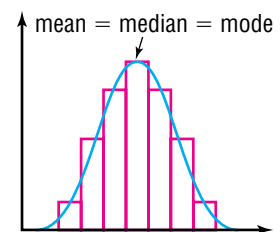
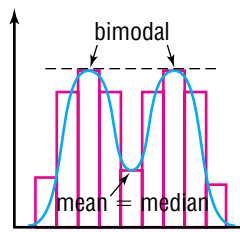
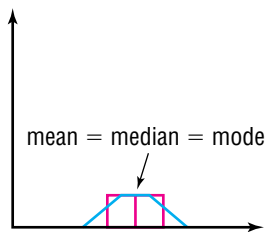
- Use the normal distribution curve.



**TESTING** The class of 1996 was the first class to take the adjusted Scholastic Assessment Test. The test was adjusted so that the median of the scores for the verbal section and the math section would be 500. For each section, the lowest score is 200 and the highest is 800. Suppose the verbal and math scores follow the *normal distribution*. What percent of the students taking the test would have a math score between 375 and 625? *This problem will be solved in Example 4.*

A frequency polygon displays a limited number of data and may not represent an entire population. To display the frequency of an entire population, a smooth curve is used rather than a polygon.

If the curve is symmetric, then information about the measures of central tendency can be gathered from the graph. Study the graphs below.



A **normal distribution** is a frequency distribution that often occurs when there is a large number of values in a set of data. The graph of this distribution is a symmetric, bell-shaped curve, shown at the left. This is known as a **normal curve**. The shape of the curve indicates that the frequencies in a normal distribution are concentrated around the center portion of the distribution. A small portion of the population occurs at the extreme values.

In a normal distribution, small deviations are much more frequent than large ones. Negative deviations and positive deviations occur with the same frequency. The points on the horizontal axis represent values that are a certain number of standard deviations from the mean  $\bar{X}$ . In the curve shown above, each interval represents one standard deviation. So, the section from  $\bar{X}$  to  $\bar{X} + \sigma$  represents those values between the mean and one standard deviation greater than the mean, the section from  $\bar{X} + \sigma$  to  $\bar{X} + 2\sigma$  represents the interval one standard deviation greater than the mean to two standard deviations greater than the mean, and so on. The total area under the normal curve and above the horizontal axis represents the total probability of the distribution, which is 1.

**Example**



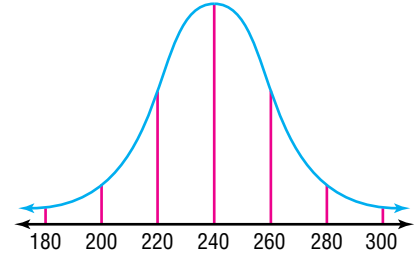
**1 MEDICINE** The average healing time of a certain type of incision is 240 hours with a standard deviation of 20 hours. Sketch a normal curve that represents the frequency of healing times.

First, find the values defined by the standard deviation in a normal distribution.

$$\begin{aligned}\bar{X} - 1\sigma &= 240 - 1(20) \text{ or } 220 \\ \bar{X} - 2\sigma &= 240 - 2(20) \text{ or } 200 \\ \bar{X} - 3\sigma &= 240 - 3(20) \text{ or } 180\end{aligned}$$

$$\begin{aligned}\bar{X} + 1\sigma &= 240 + 1(20) \text{ or } 260 \\ \bar{X} + 2\sigma &= 240 + 2(20) \text{ or } 280 \\ \bar{X} + 3\sigma &= 240 + 3(20) \text{ or } 300\end{aligned}$$

Sketch the general shape of a normal curve. Then, replace the horizontal scale with the values you have calculated.



The tables below give the fractional parts of a normally distributed set of data for selected areas about the mean. The letter  $t$  represents the number of standard deviations from the mean (that is,  $\bar{X} \pm t\sigma$ ). When  $t = 1$ ,  $t$  represents 1 standard deviation above and below the mean.

$P$  represents the fractional part of the data that lies in the interval  $\bar{X} \pm t\sigma$ . The percent of the data within these limits is  $100P$ .

$t$	$P$
0.0	0.000
0.1	0.080
0.2	0.159
0.3	0.236
0.4	0.311
0.5	0.383
0.6	0.451
0.7	0.516
0.8	0.576

$t$	$P$
0.9	0.632
1.0	0.683
1.1	0.729
1.2	0.770
1.3	0.807
1.4	0.838
1.5	0.866
1.6	0.891
1.65	0.900

$t$	$P$
1.7	0.911
1.8	0.929
1.9	0.943
1.96	0.950
2.0	0.955
2.1	0.964
2.2	0.972
2.3	0.979
2.4	0.984

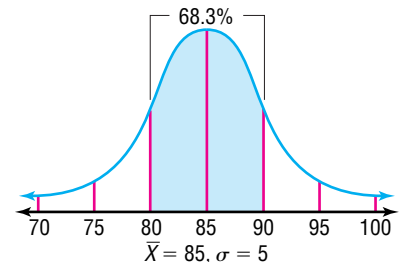
$t$	$P$
2.5	0.988
2.58	0.990
2.6	0.991
2.7	0.993
2.8	0.995
2.9	0.996
3.0	0.997
3.5	0.9995
4.0	0.9999

The  $P$  value also corresponds to the probability that a randomly selected member of the sample lies within  $t$  standard deviation units of the mean. For example, suppose the mean of a set of data is 85 and the standard deviation is 5.

**Boundaries:**

$$\begin{aligned}\bar{X} - t\sigma &\text{ to } \bar{X} + t\sigma \\ 85 - t(5) &\text{ to } 85 + t(5) \\ 85 - 1(5) &\text{ to } 85 + 1(5) \\ &80 \text{ to } 90\end{aligned}$$

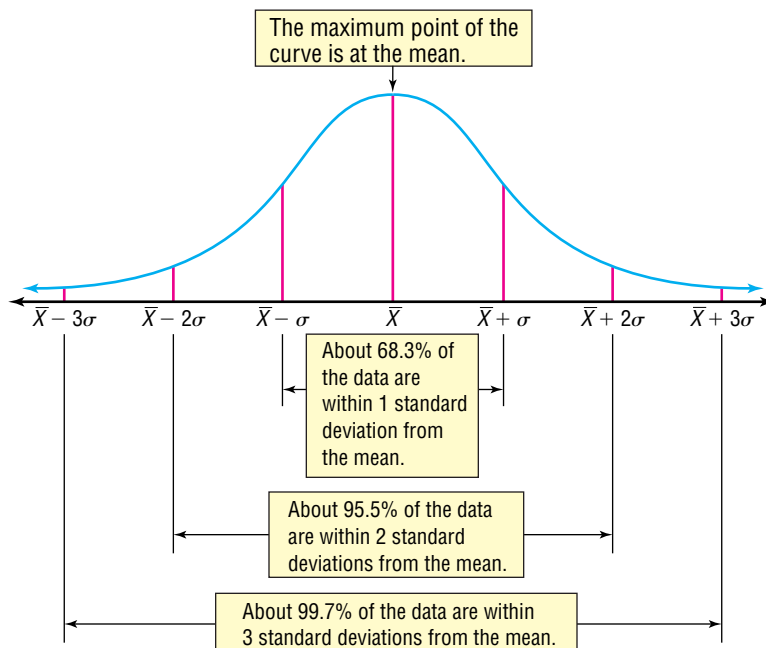
68.3% of the values in this set of data lie within one standard deviation of 85; that is, between 80 and 90.





If you randomly select one item from the sample, the probability that the one you pick will be between 80 and 90 is 0.683. If you repeat the process 1000 times, approximately 68.3% (about 683) of those selected will be between 80 and 90.

Thus, normal distributions have the following properties.



**Example**



**2 MEDICINE** Refer to Example 1. Suppose a hospital has treated 2000 patients in the past five years having this type of incision. Estimate how many patients healed in each of the following intervals.

**a. 220–260 hours**

The interval 220–260 hours represents  $\bar{X} \pm 1\sigma$ , which represents a probability of 68.3%.

$$68.3\%(2000) = 1366$$

Approximately 1366 patients took between 220 and 260 hours to heal.

**b. 200–280 hours**

The interval 200–280 hours represents  $\bar{X} \pm 2\sigma$ , which represents a probability of 95.5%.

$$95.5\%(2000) = 1910$$

Approximately 1910 patients took between 200 and 280 hours to heal.

**c. 180–300 hours**

The interval 180–300 hours represents  $\bar{X} \pm 3\sigma$ , which represents a probability of 99.7%.

$$99.7\%(2000) = 1994$$

Approximately 1994 patients took between 180 and 300 hours to heal.



If you know the mean and the standard deviation, you can find a range of values for a given probability.

**Example 3** Find the upper and lower limits of an interval about the mean within which 45% of the values of a set of normally distributed data can be found if  $\bar{X} = 110$  and  $\sigma = 15$ .

Use the table on page 919 to find the value of  $t$  that most closely approximates  $P = 0.45$ . For  $t = 0.6$ ,  $P = 0.451$ . Choose  $t = 0.6$ . Now find the limits.

$$\begin{aligned}\bar{X} \pm t\sigma &= 110 \pm 0.6(15) & \bar{X} &= 110, t = 0.6, \sigma = 15 \\ &= 101 \text{ and } 119\end{aligned}$$

The interval in which 45% of the data lies is 101–119.

If you know the mean and standard deviation, you can also find the percent of the data that lies within a given range of values.

**Example 4 TESTING** Refer to the application at the beginning of the lesson.



a. Determine the standard deviation.

b. What percent of the students taking the test would have a math score between 375 and 625?

c. What is the probability that a senior chosen at random has a math score between 550 and 650?

a. For a normal distribution, both the median and the mean are 500. All of the scores are between 200 and 800. Therefore, all of the scores must be within 300 points from the mean. In a normal distribution, 0.9999 of the data is within 4 standard deviations of the mean. If the scores are to be a normal distribution, the standard deviation should be  $\frac{300}{4}$  or 75.

b. Write each of the limits in terms of the mean.

$$375 = 500 - 125 \text{ and } 625 = 500 + 125$$

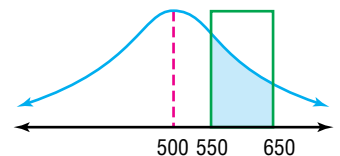
Therefore,  $\bar{X} \pm t\sigma = 500 \pm 125$  and  $t\sigma = 125$ . Solve for  $t$ .

$$\begin{aligned}t\sigma &= 125 \\ t(75) &= 125 & \sigma &= 75 \\ t &\approx 1.7\end{aligned}$$

If  $t = 1.7$ , then  $P = 0.911$ . Use the table on page 919.

About 91.1% of the students taking the test would have a math score between 375 and 625.

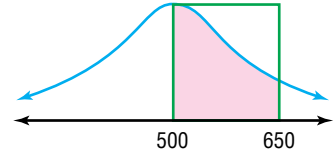
c. The graph shows that 550–650 does not define an interval that can be represented by  $\bar{X} \pm t\sigma$ . However, the interval can be defined as the difference between the intervals 500–650 and 500–550.



(continued on the next page)

First, find the probability that the score is between the mean 500 and the upper limit 650.

$$\begin{aligned}\bar{X} + t\sigma &= 650 \\ 500 + t(75) &= 650 \\ t &= 2\end{aligned}$$

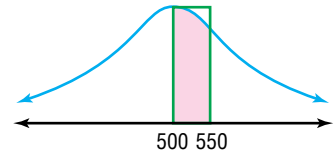


The value of  $P$  that corresponds to  $t = 2$  is 0.955.

$P = 0.955$  describes the probability that a student's score falls  $\pm 2(75)$  points about the mean, or between 350 and 650, but we are only considering half that interval. So, the probability that a student's score is between 500 and 650 is  $\frac{1}{2}(0.955)$  or about 0.478.

Next, find the probability that a score is between the mean and the lower limit 550.

$$\begin{aligned}\bar{X} + t\sigma &= 550 \\ 500 + t(75) &= 550 \\ t &\approx 0.7\end{aligned}$$

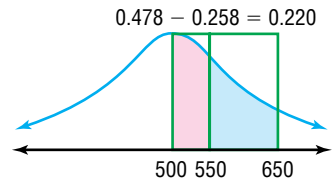


For  $t = 0.7$ ,  $P = 0.516$ .

Likewise, we will only consider half of this probability or 0.258.

Now find the probability that a student's score falls in the interval 550–650.

$$\begin{aligned}P(550-650) &= P(500-650) - P(500-550) \\ P &\approx 0.478 - 0.258 \\ P &\approx 0.220 \text{ or } 22\%\end{aligned}$$



The probability that a student's score is between 550 and 650 is about 22%.

Students who take the SAT or ACT tests will receive a score as well as a **percentile**. The percentile indicates how the student's score compares with other students taking the test.

### Percentile

The  $n$ th percentile of a set of data is the value in the set such that  $n$  percent of the data is less than or equal to that value.

Therefore if a student scores in the 65th percentile, this means that 65% of the students taking the test scored the same or less than that student.

## CHECK FOR UNDERSTANDING

### Communicating Mathematics

Read and study the lesson to answer each question.

- Compare** the median, mean, and mode of a set of normally distributed data.
- Write** an expression for the interval that is within 1.5 standard deviations from the mean.
- Sketch** a normal curve with a mean of 75 and a standard deviation of 10 and a normal curve with a mean of 75 and a standard deviation of 5. Which curve displays less variability?

4. **Counterexample** Draw a curve that represents data which is not normally distributed.
5. **Name** the percentile that describes the median.

**Guided Practice**

6. The mean of a set of normally distributed data is 550 and the standard deviation is 35.
- Sketch a curve that represents the frequency distribution.
  - What percent of the data is between 515 and 585?
  - Name the interval about the mean in which about 99.7% of the data are located.
  - If there are 200 values in the set of data, how many would be between 480 and 620?
7. A set of 500 values is normally distributed with a mean of 24 and a standard deviation of 2.
- What percent of the data is in the interval 22-26?
  - What percent of the data is in the interval 20.5-27.5?
  - Find the interval about the mean that includes 50% of the data.
  - Find the interval about the mean that includes 95% of the data.
8. **Education** In her first semester of college, Salali earned a grade of 82 in chemistry and a grade of 90 in speech.
- The mean of the chemistry grades was 73, and the standard deviation was 3. Draw a normal distribution for the chemistry grades.
  - The mean of the speech grades was 80, and the standard deviation was 5. Draw a normal distribution for the speech grades.
  - Which of Salali's grades is relatively better based on standard deviation from the mean? Explain.

## EXERCISES

**Practice**

9. The mean of a set of normally distributed data is 12 and the standard deviation is 1.5.
- Sketch a curve that represents the frequency distribution.
  - Name the interval about the mean in which about 68.3% of the data are located.
  - What percent of the data is between 7.5 and 16.5?
  - What percent of the data is between 9 and 15?
10. Suppose 200 values in a set of data are normally distributed.
- How many values are within one standard deviation of the mean?
  - How many values are within two standard deviations of the mean?
  - How many values fall in the interval between the mean and one standard deviation above the mean?
11. A set of data is normally distributed with a mean of 82 and a standard deviation of 4.
- Find the interval about the mean that includes 45% of the data.
  - Find the interval about the mean that includes 80% of the data.
  - What percent of the data is between 76 and 88?
  - What percent of the data is between 80.5 and 83.5?



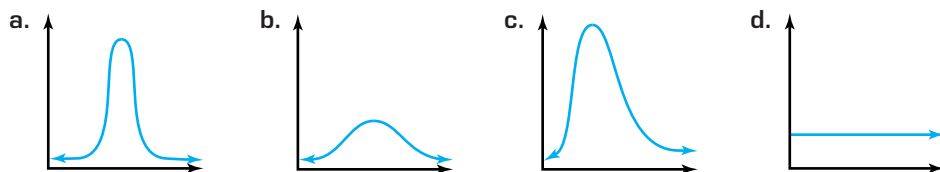
12. The mean of a set of normally distributed data is 402, and the standard deviation is 36.
- Find the interval about the mean that includes 25% of the data.
  - What percent of the data is between 387 and 417?
  - What percent of the data is between 362 and 442?
  - Find the interval about the mean that includes 45% of the data.
13. A set of data is normally distributed with a mean of 140 and a standard deviation of 20.
- What percent of the data is between 100 and 150?
  - What percent of the data is between 150 and 180?
  - Find the value that defines the 75th percentile.
14. The mean of a set of normally distributed data is 6, and the standard deviation is 0.35.
- What percent of the data is between 6.5 and 7?
  - What percent of the data is between 5.5 and 6.2?
  - Find the limit above which 90% of the data lies.
15. **Probability** Tossing six coins is a binomial experiment.
- Find each probability.  

$P(\text{no tails})$	$P(\text{one tail})$	$P(\text{two tails})$	$P(\text{three tails})$
$P(\text{four tails})$	$P(\text{five tails})$	$P(\text{six tails})$	
  - Assume that the experiment was repeated 64 times. Make a bar graph showing how many times you would expect each outcome to occur.
  - Use the bar graph to determine the mean number of tails.
  - Find the standard deviation of the number of tails.
  - Compare the bar graph to a normal distribution.
16. **Critical Thinking** Consider the percentile scores on a standardized test.
- Describe the 92nd percentile in terms of standard deviation.
  - Name the percentile of a student whose score is 0.8 standard deviation above the mean.
17. **Health** The lengths of babies born in City Hospital in the last year are normally distributed. The mean length is 20.4 inches, and the standard deviation is 0.8 inch. Trey was 22.3 inches long at birth.
- What percent of the babies born at City Hospital were longer than Trey?
  - What percent of the babies born at City Hospital were shorter than Trey?
18. **Business** The length of time a brand of CD players can be used before needing service is normally distributed. The mean length of time is 61 months, and the standard deviation is 5 months. The manufacturer plans to issue a guarantee that it will replace any CD player that breaks within a certain length of time. If the manufacturer does not want to replace any more than 2% of the CD players, how many months should they limit the guarantee?
19. **Education** A college professor plans to grade a test on a curve. The mean score on the test is 65, and the standard deviation is 7. The professor wants 15% A's, 20% B's, 30% C's, 20% D's and 15% F's. Assume the grades are normally distributed.
- What is the lowest score for an A?
  - Find the lowest passing score.
  - What is the interval for the B's?

**Applications  
and Problem  
Solving**



20. **Critical Thinking** Describe the frequency distribution represented by each graph.



21. **Industry** A machine is used to fill cans of cola. The amount of cola dispensed into each can varies slightly. Suppose the amount of cola dispensed into the cans is normally distributed.

- If at least 95% of the cans must have between 350 and 360 milliliters of cola, find the greatest standard deviation that can be allowed.
- What percent of the cans will have between 353 and 357 milliliters of cola?

**Mixed Review**

22. **Nutrition** The numbers of Calories in one serving of twenty different cereals are listed below. (*Lesson 14-3*)

110, 110, 330, 200, 88, 110, 88, 110, 165, 390,  
150, 440, 536, 200, 110, 165, 88, 147, 110, 165

- What is the median of the data?
- Find the first quartile point and the third quartile point.
- Find the interquartile range of the data.
- What is the semi-quartile range of the data?
- Draw a box-and-whisker plot of the data.



23. Find the mean, median, and mode of {33, 42, 71, 19, 42, 45, 79, 48, 55}. (*Lesson 14-2*)

24. Write an equation for a secant function with a period of  $\frac{\pi}{2}$ , a phase shift of  $-\pi$ , and a vertical shift of 3. (*Lesson 6-7*)

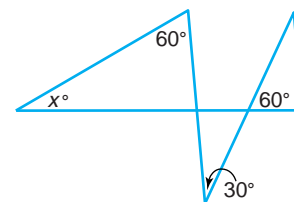
25. **Education** The numbers of students attending Wilder High School during various years are listed below. (*Lesson 4-8*)

Year	1965	1970	1975	1980	1985	1990	1995	2000
Enrollment	365	458	512	468	426	401	556	629

- Write an equation that models the enrollment as a function of the number of years since 1965.
- Use the model to predict the enrollment in 2015.

26. **SAT/ACT Practice** What is the value of  $x$  in the figure at the right?

- A 50                      B 45  
C 40                      D 35  
E 30





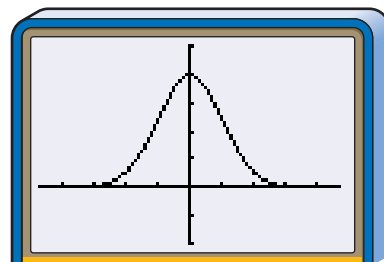
# 14-4B The Standard Normal Curve

An Extension of Lesson 14-4

## WHAT YOU'LL LEARN

- Use the standard normal curve to study properties of normal distributions.

The graph shown at the right is known as the *standard normal curve*. The standard normal curve is the graph of  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ . You can use a graphing calculator to investigate properties of this function and its graph. Enter the function for the normal curve in the **Y=** list of a graphing calculator.



**[-4.7, 4.7] scl:1 by [-0.2, 0.5] scl:0.1**

## TRY THESE

- What can you say about the function and its graph? Be sure to include information about the domain, the range, symmetry, and end behavior.
- The standard normal curve models a probability distribution. As a result, probabilities for intervals of  $x$ -values are equal to areas of regions bounded by the curve, the  $x$ -axis, and the vertical lines through the endpoints of the intervals. The calculator can approximate the areas of such regions. To find the area of the region bounded by the curve, the  $x$ -axis, and the vertical lines  $x = -1$  and  $x = 1$ , go to the **CALC** menu and select **7:∫f(x) dx**. Move the cursor to the point where  $x = -1$ . Press **ENTER**. Then move the cursor to the point where  $x = 1$  and press **ENTER**. The calculator will shade the region described above and display its approximate area. What number does the calculator display for the area of the shaded region?
- Refer to the diagram on page 920. For normal distributions, about what percent of the data are within one standard deviation from the mean? How is this number related to the area you found in Exercise 2?
- Enter **2nd** **[DRAW]** **1**. This causes the calculator to clear the shading and redisplay the graph. Find the area of the region bounded by the curve, the  $x$ -axis, and the vertical lines  $x = -2$  and  $x = 2$ .
- Find the area of the region bounded by the curve, the  $x$ -axis, and the vertical lines  $x = -3$  and  $x = 3$ .
- How do your answers for Exercises 4 and 5 compare to the percents in the diagram on page 920?
- Without using a calculator, estimate the area of the region bounded by the curve, the  $x$ -axis, and the vertical lines  $x = -4$  and  $x = 4$  to four decimal places.
- Change the graphing window to **Xmin** =  $-47$  and **Xmax** =  $47$ . Find the area of the region bounded by the curve, the  $x$ -axis, and the vertical lines  $x = -20$  and  $x = 20$ . Do you think that your answer is the exact area for the region? Explain.

## WHAT DO YOU THINK?



# Sample Sets of Data

## OBJECTIVE

- Find the standard error of the mean to predict the true mean of a population with a certain level of confidence.



**EDUCATION** Rosalinda Perez is doing some research for her doctoral thesis. She wants to determine the mean amount of money spent by a school district to educate one student for a year. Since there are 14,883 school districts in the United States, she cannot contact every school district. She randomly contacts 100 of these districts to find out how much money they spend per pupil. Using these 100 values, she computes the mean expenditure to be \$6130 with a standard deviation of \$1410. What is the standard error of the mean? *This problem will be solved in Example 1.*



In statistics, the word **population** does not always refer to a group of people. A population is the entire set of items or individuals in the group being considered. In the application above, the population is the 14,883 school districts. Rarely will a researcher find 100% of a population accessible as a source of data. Therefore, a **random sample** of the population is selected so that it is representative of the entire population. A random sample will give each member of the population the same chance of being selected as a member of the sample.

If the sample is random, the characteristics of the population pertinent to the study should be found in the sample in about the same ratio as they exist in the total population. If Ms. Perez randomly selected the school districts in the sample, the sample will probably include some large school districts, some small school districts, some urban school districts, and some rural school districts. It would probably also include school districts from all parts of the United States.

Based upon a random sample of the population, certain inferences can be made about the population in general. The major purpose of **inferential statistics** is to use the information gathered in a sample to make predictions about a population.

Ms. Perez is not sure that the mean school district expenditure per student is a truly representative mean of all school districts. She uses another sample of 100 districts. This time she finds the mean of the per pupil expenditures to be \$6080 with a standard deviation of \$1390.

The discrepancies that Ms. Perez found in her two samples are common when taking random samples. Large companies and statistical organizations often take many samples to find the “average” they are seeking. For this reason, a sample mean is assumed to be near its true population mean, symbolized by  $\mu$ . The standard deviation of the distribution of the sample means is known as the **standard error of the mean**.

$\mu$  is the lowercase Greek letter mu.



## Standard Error of the Mean

If a sample of data has  $N$  values and  $\sigma$  is the standard deviation, the standard error of the mean  $\sigma_{\bar{x}}$  is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}.$$

*The symbol  $\sigma_{\bar{x}}$  is read “sigma sub  $x$  bar.”*

The standard error of the mean is a measurement of how well the mean of a sample selected at random estimates the true mean.

- If the standard error of the mean is small, then the sample means are closer to (or approximately the same value as) the true mean.
- If the standard error of the mean is large, many of the sample means would be far from the true mean.

The greater the number of items or subjects in the sample, the closer the sample mean reflects the true mean and the smaller the standard error of the mean.

### Example



#### 1 EDUCATION Refer to the application at the beginning of the lesson. What is the standard error of the mean?

For the sample,  $N = 100$ . Find  $\sigma_{\bar{x}}$ .

$$\sigma_{\bar{x}} = \frac{1410}{\sqrt{100}} \text{ or } 141$$

The standard error of the mean of the per pupil expenditures is 141. Although 141 is a fairly large number, it is not extremely large compared to the mean 6130. The mean is a good approximation of the true mean.

*The sampling error is the difference between the population mean  $\mu$  and the sample mean  $\bar{X}$ .*

The sample mean is only an estimate of the true mean of the population. Sample means of various random samples of the same population are normally distributed about the true mean with the standard error of the mean as a measure of their variability. Thus, the standard error of the mean behaves like the standard deviation. Using the standard error of the mean and the sample mean, we can state a range about the sample mean in which we think the true mean lies. Probabilities of the occurrence of sample means and true means may be determined by referring to the tables on page 919.

### Example



#### 2 EDUCATION Refer to the application at the beginning of the lesson. Using Ms. Perez’s first sample, determine the interval of per pupil expenditures such that the probability is 95% that the mean expenditure of the entire population lies within the interval.

When  $P = 95\%$  or  $0.95$ ,  $t = 1.96$ . *Refer to the tables on page 919.*

To find the range, use a technique similar to finding the interval for a normal distribution.

$$\begin{aligned}\bar{X} \pm t\sigma_{\bar{x}} &= 6130 \pm 1.96(141) \\ &= 5853.64 \text{ to } 6406.36\end{aligned}$$

The probability is 95% that the true mean  $\mu$  is within the interval of \$5853.64 and \$6406.36.

The probability of the true mean being within a certain range of a sample mean may be expressed as a **level of confidence**. The most commonly used levels of confidence are 1% and 5%. A 1% level of confidence means that there is less than a 1% chance that the true mean differs from the sample mean by a certain amount. That is, you are 99% confident that the true mean is within a certain range of the sample mean. A 5% level of confidence means that the probability of the true mean being within a certain range of the sample mean is 95%.

If a higher level of confidence is desired for the same number of values, accuracy must be sacrificed by providing a larger interval. However, if the number of values in the sample is larger, the interval for a given level of confidence is smaller.

**Example**



**3 AUTOMOTIVE ENGINEERING** The number of miles a certain sport utility vehicle can travel on open highway on one gallon of gasoline is normally distributed. You are to take a sample of vehicles, test them, and record the miles per gallon. You wish to have a 1% level of confidence that the interval containing the mean miles per gallon of the sample also contains the true mean.

- a. Twenty-five sports utility vehicles are selected and tested. From this sample, the average miles per gallon is 22 with a standard deviation of 4 miles per gallon. Determine the interval about the sample mean that has a 1% level of confidence.
- b. Four hundred sports utility vehicles are randomly selected and their miles per gallon are recorded. From this sample, the average miles per gallon is 22 with a standard deviation of 4 miles per gallon. Determine the interval about the sample mean that has a 1% level of confidence.
- c. What happens when the number of items in the sample is increased?

- a. A 1% level of confidence is given when  $P = 99\%$ .  
When  $P = 0.99$ ,  $t = 2.58$ .

$$\text{Find } \sigma_{\bar{X}} \quad \sigma_{\bar{X}} = \frac{4}{\sqrt{25}} \text{ or } 0.8$$

$$\text{Find the range.} \quad \begin{aligned} \bar{X} \pm t\sigma_{\bar{X}} &= 22 \pm 2.58(0.8) \\ \bar{X} \pm t\sigma_{\bar{X}} &= 19.936 \text{ and } 24.064 \end{aligned}$$

Thus, the interval about the sample mean is 19.936 miles per gallon to 24.064 miles per gallon.

- b. Determine the range. As in part a,  $t = 2.58$ .

$$\text{Find } \sigma_{\bar{X}}. \quad \sigma_{\bar{X}} = \frac{4}{\sqrt{400}} \text{ or } 0.2$$

$$\text{Find the range.} \quad \begin{aligned} \bar{X} \pm t\sigma_{\bar{X}} &= 22 \pm 2.58(0.2) \\ \bar{X} \pm t\sigma_{\bar{X}} &= 21.484 \text{ and } 22.516 \end{aligned}$$

Thus, the interval about the sample mean is 21.484 miles per gallon to 22.516 miles per gallon.

- c. By increasing the number of items in the sample, but with the same 1% confidence level, the range decreased substantially, from about 24.064 – 19.936 or 4.128 miles per gallon to about 22.516 – 21.484 or 1.032 miles per gallon.



## CHECK FOR UNDERSTANDING

### Communicating Mathematics

Read and study the lesson to answer each question.

1. **Compare and contrast** a population and a sample of the population.
2. **Describe** how to determine the standard error of the mean.
3. **Explain** how you can find a smaller interval for the true mean and still have the same level of confidence.
4. **You Decide** As part of a class project, Karen, Tyler, and Mark need to determine the average number of hours that the students in their high school study each school night. Karen suggests they ask the students in their senior English class. Tyler suggests they ask every twentieth student as they enter the school in the morning. Mark suggests they ask the members of the track team. Whose sample should they use? Explain.

### Guided Practice

Find the standard error of the mean for each sample.

5.  $\sigma = 73, N = 100$
6.  $\sigma = 3.4, N = 250$
7. If  $\sigma = 5, N = 36$ , and  $\bar{X} = 45$ , find the interval about the sample mean that has a 1% level of confidence.
8. If  $\sigma = 5.6, N = 300$ , and  $\bar{X} = 55$ , find the interval about the sample mean that has a 5% level of confidence.
9. **Employment** An employment agency requires all clients to take an aptitude test. They randomly selected 150 clients and recorded the amount of time each client took to complete the test. The average time needed to complete the test was 27.5 minutes with a standard deviation of 3.5 minutes.
  - a. What is the standard error of the mean?
  - b. Find the interval about the sample mean that reflects a 50% chance that the true mean lies within that interval.
  - c. Find the interval about the sample mean that has a 1% level of confidence.

## EXERCISES

### Practice

Find the standard error of the mean for each sample.

10.  $\sigma = 1.8, N = 81$
11.  $\sigma = 5.8, N = 250$
12.  $\sigma = 7.8, N = 140$
13.  $\sigma = 14, N = 700$
14.  $\sigma = 2.7, N = 130$
15.  $\sigma = 13.5, N = 375$
16. If the standard deviation of a sample set of data is 5.6 and the standard error of the mean is 0.056, how many values are in the sample set?

For each sample, find the interval about the sample mean that has a 1% level of confidence.

17.  $\sigma = 5.3, N = 50, \bar{X} = 335$
18.  $\sigma = 40, N = 64, \bar{X} = 200$
19.  $\sigma = 12, N = 200, \bar{X} = 80$
20.  $\sigma = 11.12, N = 1000, \bar{X} = 110$

21. The mean height of a sample of 100 high school seniors is 68 inches with a standard deviation of 4 inches. Determine the interval of heights such that the probability is 90% that the mean height of the entire population lies within that interval.

For each sample, find the interval about the sample mean that has a 5% level of confidence.

22.  $\sigma = 2.4, N = 100, \bar{X} = 24$                       23.  $\sigma = 17.1, N = 350, \bar{X} = 4526$   
 24.  $\sigma = 28, N = 370, \bar{X} = 678$                       25.  $\sigma = 0.67, N = 80, \bar{X} = 5.38$

26. The following is a frequency distribution of the time in minutes required for a shopper to get through the checkout line at a certain discount store on a weekend. The distribution is a random sample of the weekend shoppers at the store.

Number of Minutes	3-5	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21
Frequency	1	3	5	12	17	13	7	4	2

- What is the mean of the data in the frequency distribution?
  - Find the standard deviation of the data.
  - What is the standard error of the mean?
  - Find the interval about the sample mean such that the probability is 0.95 that the true mean lies within the interval.
  - Find the probability that the mean of the population will be less than one minute from the mean of the sample.
27. The standard deviation of the blood pressure of 45 women ages 40 to 50 years old is 12. What is the probability that the mean blood pressure of the random sample will differ by more than 3 points from the mean blood pressure reading for all women in that age bracket?

**Applications  
and Problem  
Solving**



28. **Botany** A botanist is studying the effects of a drought on the size of acorns produced by the oak trees. A random sample of 50 acorns reveals a mean diameter of 16.2 millimeters and a standard deviation of 1.4 millimeters.
- Find the standard error of the mean.
  - What is the interval about the sample mean that has a 5% level of confidence?
  - Find the interval about the sample mean that gives a 99% chance that the true mean lies within the interval.
  - Find the interval about the sample mean such that the probability is 0.80 that the true mean lies within the interval.
29. **Advertising** The Brite Light Company wishes to include the average lifetime of its light bulbs in its advertising. One hundred light bulbs are randomly selected and illuminated. The time for each bulb to burn out is recorded. From this sample, the average life is 350 hours with a standard deviation of 45 hours.
- Find the standard error of the mean.
  - Determine the interval about the sample mean that has a 1% level of confidence.
  - If you want to avoid false advertising charges, what number would you use as the average lifetime of the bulbs? Explain.

- 30. Critical Thinking** There is a probability of 0.99 that the average life of a disposable hand-warming package is between 9.7936 and 10.2064 hours. The standard deviation of the sample is 0.8 hour. What is the size of the sample used to determine these values?
- 31. Entertainment** In a certain town, a random sample of 10 families was interviewed about their television viewing habits. The mean number of hours that these families had their televisions on per day was 4.1 hours. The standard deviation was 1.8 hours.
- What is the standard error of the mean?
  - Make a 5% level of confidence statement about the mean number of hours that families of this town have their televisions turned on per day.
  - What inferences about the television habits of the entire city can be drawn from this data? Explain.
- 32. Quality Control** The Simply Crackers Company selects a random sample of 50 snack packages of their cheese crackers. The mean number of crackers in a package is 42.7 with a standard deviation of 3.2.
- Find the standard error of the mean.
  - Determine the interval of number of crackers such that the probability is 50% that the mean number of crackers lies within the interval.
  - If the true population mean should be 43 crackers per package, should the company be concerned about this sample? Explain.
- 33. Critical Thinking** The lifetimes of 1600 batteries used in radios are tested. With a 5% level of confidence, the true average life of the batteries is from 746.864 to 753.136 hours.
- What is the mean life of a battery in the sample?
  - Find the standard deviation of the life of the batteries in the sample.

### Mixed Review

- 34. Tires** The lifetimes of a certain type of car tire are normally distributed. The mean lifetime is 40,000 miles with a standard deviation of 5000 miles. Consider a sample of 10,000 tires. (*Lesson 14-4*)
- How many tires would you expect to last between 35,000 and 45,000 miles?
  - How many tires would you expect to last between 30,000 and 40,000 miles?
  - How many tires would you expect to last less than 40,000 miles?
  - How many tires would you expect to last more than 50,000 miles?
  - How many tires would you expect to last less than 25,000 miles?
- 35.** Find the mean deviation and the standard deviation of {44, 72, 58, 61, 71, 49, 55, 68}. (*Lesson 14-3*)
- 36.** Find the sum of the first ten terms of the series  $\frac{1}{16} + \frac{1}{4} + 1 + \dots$ . (*Lesson 12-2*)
- 37.** Write  $x = y$  in polar form. (*Lesson 9-3*)
- 38.** Solve  $\tan x + \cot x = 2$  for principal values of  $x$ . (*Lesson 7-5*)
- 39. SAT/ACT Practice**  $*x$  is defined such that  $*x = x^2 - 2x$ . What is the value of  $*2 - *1$ ?

A -1

B 0

C 1

D 2

E 4

## VOCABULARY

arithmetic mean (p. 897)

back-to-back bar graph  
(p. 889)

bar graph (p. 889)

bimodal (p. 899)

box-and-whisker plot (p. 909)

class interval (p. 890)

class limits (p. 890)

class mark (p. 890)

cumulative frequency  
distribution (p. 902)

frequency distribution  
(p. 890)

frequency polygon (p. 892)

hinge (p. 909)

histogram (p. 890)

inferential statistics (p. 927)

interquartile range (p. 909)

leaf (p. 899)

level of confidence (p. 929)

line plot (p. 889)

mean (p. 897)

mean deviation (p. 910)

measure of central tendency  
(p. 897)

measure of variability  
(p. 908)

median (p. 897)

median class (p. 902)

mode (p. 897)

normal curve (p. 918)

normal distribution (p. 918)

outlier (p. 909)

percentile (p. 922)

population (p. 927)

quartile (p. 909)

random sample (p. 927)

range (p. 890)

semi-interquartile range  
(p. 909)

standard deviation (p. 911)

standard error of the mean  
(p. 927)

stem (p. 899)

stem-and-leaf plot (p. 899)

three-dimensional bar graph  
(p. 890)

variance (p. 912)

whisker (p. 909)

## UNDERSTANDING AND USING THE VOCABULARY

Choose the term from the list above that best completes each statement.

1. A \_\_\_\_\_ is a display that visually shows the quartile points and the extreme values of a set of data.
2. The \_\_\_\_\_ of a set of data is the middle value if there are an odd number of values.
3. The standard deviation of the distribution of the sample means is known as the \_\_\_\_\_.
4. The \_\_\_\_\_ of a set of data is the difference between the greatest and the least values in the set.
5. A statistic that describes the center of a set of data is called an average or \_\_\_\_\_.
6. A \_\_\_\_\_ is the entire set of items or individuals in the group being considered.
7. Data with two modes are \_\_\_\_\_.
8. The major purpose of \_\_\_\_\_ is to use the information gathered in a sample to make predictions about a population.
9. A \_\_\_\_\_ is the most common way of displaying a frequency distribution.
10. A measure of variability often associated with the arithmetic mean is the \_\_\_\_\_.

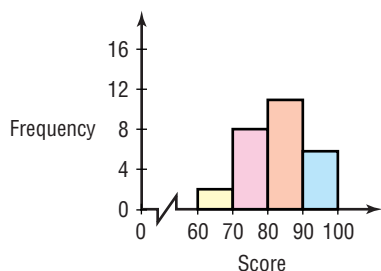
SKILLS AND CONCEPTS

OBJECTIVES AND EXAMPLES

**Lesson 14-1** Draw, analyze, and use bar graphs and histograms.

Draw a histogram of the data below.

Scores	Frequency
60–70	2
70–80	8
80–90	11
90–100	6



REVIEW EXERCISES

The table below gives the weight in ounces of the popular women's tennis shoes.

Weight (ounces)	Number of Shoes
9.0–10.0	2
10.0–11.0	18
11.0–12.0	5
12.0–13.0	2
13.0–14.0	3

- What is the range of the data?
- What are the class marks?
- Draw a histogram of the data.

**Lesson 14-2** Find the mean, median, and mode of a set of data.

Find the mean, median, and mode of the set {46, 47, 59, 49, 50, 48, 58, 56, 58, 54, 53}.

$$\frac{1}{11} \sum_{i=1}^{11} X_i = \frac{46 + 47 + \dots + 54 + 53}{11} \text{ or } 53$$

The mean is 53.

To find the median, order the data.

46, 47, 48, 49, 50, 53, 54, 56, 58, 58, 64

Since there are an odd number of data, the median is the middle value. The median is 53.

The most frequent value in this set of data is 58. So, the mode is 58.

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

- {4, 8, 2, 4, 5, 5, 6, 7, 4}
- {250, 200, 160, 240, 200}
- {19, 11, 13, 15, 16}
- {6.6, 6.3, 6.8, 6.6, 6.7, 5.9, 6.4, 6.3}

stem	leaf
12	2 8
13	0 1 3 5
14	1 6
12	2 = 122



## OBJECTIVES AND EXAMPLES

**Lesson 14-3** Find the interquartile range, the semi-interquartile range, mean deviation, and standard deviation of data.

interquartile range:  $Q_3 - Q_1$

semi-interquartile range:  $Q_R = \frac{Q_3 - Q_1}{2}$

mean deviation:  $MD = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|$

standard deviation:  $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$

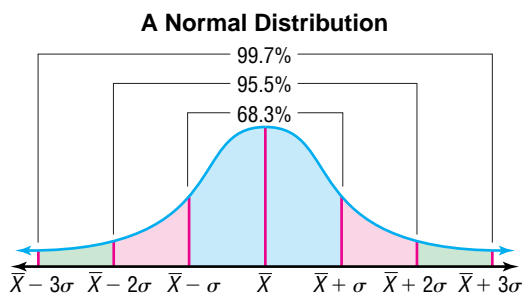
## REVIEW EXERCISES

A number cube is tossed 10 times with the following results.

5 1 5 4 2 3 6 2 5 1

19. Find the interquartile range.
20. Find the semi-interquartile range.
21. Find the mean deviation.
22. Find the standard deviation.

**Lesson 14-4** Use the normal distribution curve.



A set of data is normally distributed with a mean of 75 and a standard deviation of 6. What percent of the data is between 69 and 81?

The values within one standard deviation of the mean are between  $75 - 6$ , or 69, and  $75 + 6$ , or 81. So, 68.3% of the data is between 69 and 81.

The mean of a set of normally distributed data is 88 and the standard deviation is 5.

23. What percent of the data is in the interval 78–98?
24. Find the probability that a value selected at random from the data lies in the interval 86–90.
25. Find the interval about the mean that includes 90% of the data.

Suppose 150 values in a data set are normally distributed.

26. How many values are within one standard deviation of the mean?
27. How many values are within two standard deviations of the mean?
28. How many values fall in the interval between the mean and one standard deviation above the mean?



## OBJECTIVES AND EXAMPLES

**Lesson 14-5** Find the standard error of the mean to predict the true mean of a population with a certain level of confidence.

Find the standard error of the mean for  $\sigma = 12$  and  $N = 100$ . If  $\bar{X} = 75$ , find the range for a 1% level of confidence.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

$$\sigma_{\bar{X}} = \frac{12}{\sqrt{100}} \text{ or } 1.2$$

The standard error of the mean is 1.2.

A 1% level of confidence is given when  $P = 99\%$ .

When  $P = 0.99$ ,  $t = 2.58$ .

Use  $\sigma_{\bar{X}} = 1.2$  to find the range.

$$\begin{aligned} \bar{X} \pm t\sigma_{\bar{X}} &= 75 \pm (2.58)(1.2) \\ &= 71.90 \text{ to } 78.10 \end{aligned}$$

Thus, the interval about the mean is 71.90 to 78.10.

## REVIEW EXERCISES

Find the standard error of the mean for each sample.

29.  $\sigma = 1.5, N = 90$

30.  $\sigma = 4.9, N = 120$

31.  $\sigma = 25, N = 400$

32.  $\sigma = 18, N = 25$

For each sample, find the interval about the sample mean that has a 1% level of confidence.

33.  $\sigma = 15, N = 50, \bar{X} = 100$

34.  $\sigma = 30, N = 15, \bar{X} = 90$

35.  $\sigma = 24, N = 200, \bar{X} = 40$

In a random sample of 200 adults, it was found that the average number of hours per week spent cleaning their home was 1.8, with a standard deviation of 0.5.

36. Find the standard error of the mean.

37. Find the range about the mean such that the probability is 0.90 that the true mean lies within the range.

38. Find the range about the sample mean that has a 5% level of confidence.

39. Find the range about the sample mean that has a 1% level of confidence.

40. **Entertainment** In a random sample of 100 families, the children watched television an average of 4.6 hours a day. The standard deviation is 1.4 hours. Find the range about the sample mean so that a probability of 0.90 exists that the true mean will lie within the range.

## APPLICATIONS AND PROBLEM SOLVING

41. **Safety** The numbers of job-related injuries at a construction site for each month of 1999 are listed below. (*Lesson 14-2*)

10	13	15	39	21	24
19	16	39	17	23	25

- Make a stem-and-leaf plot of the numbers of injuries.
- What is the mean number of the data?
- Find the median of the data.
- Find the mode of the data.

42. The height of members of the boys basketball team are normally distributed. The mean height is 75 inches, and the standard deviation is 2 inches. Randall is 80 inches tall. What percent of the boys on the basketball team are taller than Randall? (*Lesson 14-4*)



## ALTERNATIVE ASSESSMENT

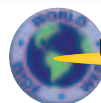
## OPEN-ENDED ASSESSMENT

- The mean of a set of five pieces of data is 15, and the median is 10. When one certain value is added to the set, the mean stays the same but the median changes.
  - Find a set of data for which this is true.
  - What value can be added to your set so that the mean stays the same but the median changes?
- Find some data in a newspaper or magazine. Use what you have learned in this chapter to analyze the data.


**PORTFOLIO**

Choose one of the types of data displays you studied in this chapter. Describe a situation in which this type of display would be used. Explain why the type of display you chose is the best one to use in this situation.

**Additional Assessment** See p. A69 for a practice Chapter 14 test.


 Unit 4 *inter*NET Project

 THE UNITED STATES  
CENSUS BUREAU

**More and more models!**

- Use the data you collected for the project in Chapter 12. Display the data in a table or use software to prepare a spreadsheet of the data.
- Use computer software or a graphing calculator to find at least three models for the population data. Draw a graph of each function.
- Compare your function models for the population data. Use your models to predict the U.S. population for the year 2050. Determine which one you think best fits the data.
- Write a one-page paper comparing the arithmetic and geometric sequences you wrote for Chapter 12 with the function models. Discuss which one model you think best fits the population data and give your estimate for the population in 2050.





## Statistics and Data Analysis Problems

On the SAT and ACT exams, you will calculate the mean (average), median, and mode of a data set.

The SAT and ACT exams may include a one or two questions on interpreting graphs. The most common graphs are bar graphs, circle graphs, line graphs, stem-and-leaf plots, histograms, and frequency tables.

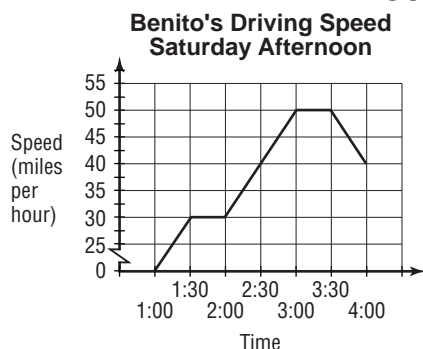
### TEST-TAKING TIP

If a problem includes a graph, look carefully at the graph including its labels and units. Then read the question.

Two or three questions may refer to the same graph.

### ACT EXAMPLES

Questions 1 and 2 refer to the following graph.



1. For what percent of the time was Benito driving 40 miles per hour or faster?
- A 20    B 25    C  $33\frac{1}{3}$     D 40    E 50

**HINT** Watch for different units of measure.

**Solution** Benito drove a total of 3 hours. He drove 40 miles per hour or faster for 1 hour and 30 minutes or  $1\frac{1}{2}$  hours. The fraction of the time he drove 40 mph or more is  $\frac{3}{2}$  or  $\frac{1}{2}$ , which equals 50%. The answer is choice E.

2. How far, in miles, did Benito drive between 1:30 and 2:00?
- A 0    B 15    C 20    D 30  
E It cannot be determined from the information given.

**Solution** rate  $\times$  time = distance  
 $(30 \text{ mph})\left(\frac{1}{2} \text{ hour}\right) = 15 \text{ miles}$

The answer is choice B.

### SAT EXAMPLE

3. For  $x = 0$ ,  $x = 1$ , and  $x = 2$ , Set  $A = \{x, x + 3, 3x, x^2\}$ . What is the mode of Set  $A$ ?

A 0    B 1    C 2    D 2.5    E 3

**HINT** Look carefully at the given information and at the form of the answer choices (numbers, variables, and so on.)

**Solution** Notice that the answer choices are numbers. But Set  $A$  is defined using variable expressions. First determine the actual data of set  $A$ . Consider each value of  $x$ , one at a time. Substitute the value for  $x$  into each element of Set  $A$ .

For  $x = 0$ :  $x = 0$ ,  $x + 3 = 3$ ,  $3x = 0$ , and  $x^2 = 0$ .

When  $x = 0$ ,  $A = \{0, 3, 0, 0\}$ .

When  $x = 1$ ,  $A = \{1, 4, 3, 1\}$ .

When  $x = 2$ ,  $A = \{2, 5, 6, 4\}$ .

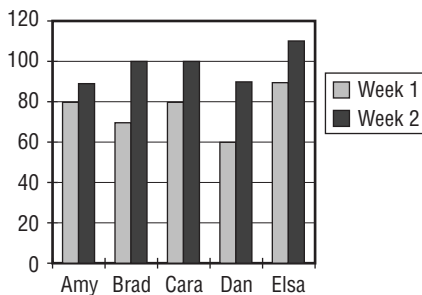
Thus,  $A = \{0, 0, 0, 1, 1, 2, 3, 3, 4, 4, 5, 6\}$ . The element 0 occurs three times and no other element occurs as many times. So the mode of Set  $A$  is 0. The answer is choice A.

You might notice that choice D, 2.5, is the value of the median set  $A$ .

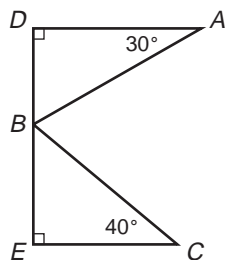
After you work each problem, record your answer on the answer sheet provided or on a piece of paper.

**Multiple Choice**

1. Based on the graph below, which worker had the greatest percent increase in income from week 1 to week 2?



- A Amy                      B Brad                      C Cara  
 D Dan                      E Elsa
2. If  $a = b + bc$ , then in terms of  $c$ , what does  $\frac{b}{a}$  equal?  
 A  $\frac{1}{c+2}$                   B  $\frac{1}{c+1}$                   C  $\frac{1}{c}$   
 D  $c$                           E  $c+1$
3. If 0.1% of  $m$  is equal to 10% of  $n$ , then  $m$  is what percent of  $10n$ ?  
 A 1/1000%                B 10%                      C 100%  
 D 1000%                  E 10,000%
4.  $S$  is the set of all positive numbers  $n$  such that  $n < 100$  and  $\sqrt{n}$  is an integer. What is the median value of the members of set  $S$ ?  
 A 5                          B 5.5                      C 25  
 D 50                        E 99
5. In the figure,  $D$ ,  $B$ , and  $E$  are collinear. What is the measure of  $\angle ABC$ ?

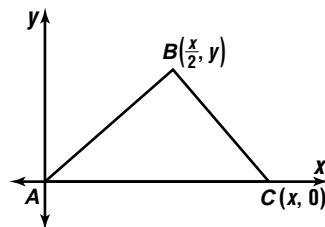


- A 20°    B 35°    C 50°    D 60°    E 70°

6. How many of the scores 10, 20, 30, 35, 35, and 50 are greater than the arithmetic mean of the scores?

- A 0                      B 1                      C 2  
 D 3                      E 4

7. In  $\triangle ABC$ , what is the ratio  $\frac{\tan A}{\text{area } \triangle ABC}$ ?



- A  $\frac{1}{2y^2}$     B  $\frac{1}{y^2}$     C  $\frac{2}{x^2}$     D  $\frac{4}{x^2}$     E  $\frac{x^2}{4}$

8. Based on the data in the table below, how many employees can this company expect to have by 2003?

Year	1997	1998	1999	2000	2001
Number of Employees	1900	2200	2500	2800	3100

- A 3100                      B 3400                      C 3550  
 D 3700                      E 4000

9. What is the difference between the median of Set  $A$  and the mean of Set  $B$ ?

Set  $A$ : {2, -1, 7, -4, 11, 3}  
 Set  $B$ : {10, 5, -3, 4, 7, -8}

- A 2  
 B 1.5  
 C 0  
 D -0.5  
 E -2

10. **Grid-In** What is the arithmetic mean of the ten numbers below?

-820, -65, -32, 0, 1, 2, 3, 32, 65, 820

**interNET CONNECTION** SAT/ACT Practice For additional test practice questions, visit: [www.amc.glencoe.com](http://www.amc.glencoe.com)