6–3 Measures of Dispersion

LEARNING OUTCOMES

1. Find the range.
2. Find measures of relative position.
3. Find the standard deviation.

LC LEARNING CATALYTICS

Evaluate.

1. $(3.7)^2$
2. $\sqrt{584}$ to the nearest tenth.

STOP AND CHECK

1. Find the range for the data given in Example 5 on page 274.
2. Find the range for the data given in Example 6 on page 274.

Answers:
1. 24 2. 18

Measures of variation or dispersion: measures that show the dispersion or spread of a set of data. Examples are the range and the standard deviation

Spread: another term for the variation or dispersion of a set of data

Range: the difference between the highest value and the lowest value in a set of data

Midrange: the mean of the lowest and highest values of a data set and is a measure of central tendency

Outliers: data items that are unusually large or unusually small when compared to the other data items

1 Find the Range. The mean, the median, and the mode are *measures of central tendency*. Other statistical measures are **measures of variation or dispersion.** The variation or dispersion of a set of data may also be referred to as the **spread.** One of these measures of dispersion is the **range.** The range is the difference between the highest value and the lowest value in a set of data.

To find the range:

1. Find the highest and lowest values.
2. Find the difference between the highest and lowest values.

EXAMPLE 1

BUS Find the range for the data described in Example 7 for fast-food restaurant hourly pay rates on page 275.

The high value is \$10.50. The low value is \$7.15.

range =

TIP **Use More Than One Statistical Measure** A common mistake when making conclusions or inferences from statistical measures is to examine only one statistic, such as the range. To obtain a complete picture of the data requires looking at more than one statistic.

In some sets of data the midrange might be a statistic of interest. The **midrange** is the mean of the lowest and highest values of the data set and is a measure of central tendency. The midrange is highly sensitive to outliers since it is determined by only two data points (highest and lowest) in the set and it is only useful when the data set has no outliers. An **outlier** is a data item that is unusually large or unusually small when compared to the other data items.

To find the midrange of a data set:

1. Find the highest and lowest values.
2. Find the mean of the highest and lowest values.

$$\text{midrange} = \frac{\text{highest value} + \text{lowest value}}{2}$$

EXAMPLE 2

Find the midrange for the data described in Example 7 for fast-food restaurant hourly pay rates on page 275.

The highest value is \$10.50. The lowest value is \$7.15.

midrange =

8.83

When data sets are compared, we often define benchmarks in the data that we will use for making the comparisons. This process is used to report scores from norm-referenced tests. The most common of these benchmarks are *quartiles* and *percentiles*.

Quartiles and percentiles are based on data that are ranked (arranged in order of size) in ascending order (from smallest to largest). First, we will look at quartiles.

Quartiles: boundaries that divide the data set into four parts

Quartiles are boundaries that divide the data set into four parts. Three boundaries will divide the data into four parts. There is more than one method for finding these division boundaries, but we will use the method for finding the median of a set of data.

To divide a set of data into quartiles:

Lower quartile: lowest quarter of data bounded by the median of the lower half of the data

Upper quartile: highest quarter of data bounded by the median of the upper half of the data

1. Find the median of the data set. This divides the data into two halves (lower half and upper half).
2. Find the median of each half of the data.
3. The median of the lower half of data is the boundary for the **lower quartile**. The median for the upper half of data is the boundary for the **upper quartile**.

STOP AND CHECK

1. Find the quartiles for the data set of 16 credit-hour loads found in Example 10 on page 276.

Answer:

1. Lower quartile: 6; median: 9; upper quartile: 12

EXAMPLE 4

Find the quartiles for the set of 25 student grades found on page 275.

Arrange the scores in ascending order.

63, 68, 71, 74, 76, 76, 77, 77, 77, 79, 81, 83, 84, 87, 87, 88, 88, 89, 90, 91, 91, 93, 93, 97, 97

Arrange the scores in ascending order.
63, 68, 71, 74, 76, 76, 77, 77, 77, 79, 81, 83, 84, 87, 87, 88, 88, 89, 90, 91, 91, 93, 93, 97, 97

25 data values, odd number, so what is the median?

Find the median.

84

Find the median of the lower half of data. That is, find the median of the values below 84.

76.5

Find the median of the upper half of data that is the median of the values above 84.

90.5

TIP **Ordinal Numbers** When data items of a set of data are arranged in ascending order, the position of a data item is identified by a value called an ordinal number. An *ordinal number* is a number that identifies an ordered position. Examples of ordinal numbers are first, second, third, and so on.

Inner quartiles: the second and third quartiles of a set of data

Inner-quartile range: the difference between the upper-quartile boundary and the lower-quartile boundary

Let's look more closely at the second and third quartiles (two quartiles in the middle). These two quartiles are referred to as the **inner quartiles.** When we looked at the midrange, we said that this statistic was not very useful because it was greatly affected by extreme values or outliers. If we just look at the inner quartiles, we eliminate all the outliers. The **inner-quartile range** is the difference between the upper-quartile boundary and the lower-quartile boundary. The inner-quartile range can be a useful statistic in examining a set of data. Half of the data will be within the inner-quartile range.

To find the inner-quartile range:

1. Arrange the data in ascending order.
2. Find the boundary for the lower quartile.
3. Find the boundary for the upper quartile.
4. Find the difference between the upper-quartile boundary and the lower-quartile boundary.

   inner-quartile range = upper-quartile boundary − lower-quartile boundary

EXAMPLE 5

Score: 0 of 1 pt   7 of 10 (0 complete)   HW Sc

6.3.21

Find the quartiles for the set of total blood serum cholesterol levels recorded by a lab.

207 192 184 222 243 213 256 231 214 198
188 204 236 224 231 265 185 193 202 207

Find the interquartile range (IQR)

The median is 210.
(Type an integer or a decimal.)

The boundary of the lower quartile is 195.5.
(Type an integer or a decimal.)

The boundary of the upper quartile is 231.
(Type an integer or a decimal.)

Find Q1, median, Q3.

IQR = Q3 - Q1

A
184
185
188
192
193
198
202
204
207
207
213
214
222
224
231
231
236
243
256
265

Data sorted

20 data values

TIP **Working with an Electronic Spreadsheet** When data are entered into an electronic spreadsheet, the data can be arranged in ascending order by sorting the data on the data column. If another column is used to count the data items (Item 1, Item 2, etc.), this column will give the rank of the sorted data. Then the 7th, 20th, or 23rd value can be found at a glance.

3 Find the Standard Deviation. Although the range gives us some information about dispersion, it does not tell us whether the highest or lowest values are typical values or extreme outliers. We can get a clearer picture of the data set by examining how much each value *differs* or *deviates* from the mean.

Deviation from the mean: the difference between a specific value and the mean

The **deviation from the mean** of a data value is the difference between a specific value and the mean.

To find the deviations from the mean:

Data set: 38, 43, 45, 44.

1. Find the mean of a set of data.

$$\bar{x} = \frac{\text{sum of data values}}{\text{number of values}} = \frac{\Sigma x_i}{n} \qquad \frac{38 + 43 + 45 + 44}{4} = \frac{170}{4} = 42.5$$

2. Find the amount that each data value deviates or is different from the mean.

deviation from the mean = data value − mean = $x_i - \bar{x}$

$38 - 42.5 = -4.5$ (below the mean)
$43 - 42.5 = 0.5$ (above the mean)
$45 - 42.5 = 2.5$ (above the mean)
$44 - 42.5 = 1.5$ (above the mean)

For values smaller than the mean, the difference is represented by a *negative* number indicating the value is *below* or less than the mean. For values larger than the mean, the difference is represented by a positive number indicating the value is *above* or greater than the mean. *The absolute value of the sum of the deviations below the mean should equal the sum of the deviations above the mean.* In the example in the box, only one value is below the mean, and its deviation is −4.5. Three values are above the mean, and the sum of these deviations is $0.5 + 2.5 + 1.5 = 4.5$. We say that *the sum of all deviations from the mean is zero.* This is true for all sets of data.

We have not gained any statistical insight or new information by analyzing the sum of the deviations from the mean or even by analyzing the average of the deviations.

$$\text{average deviation} = \frac{\text{sum of deviations}}{\text{number of values}} = \frac{0}{n} = 0$$

STOP AND CHECK

1. Find the deviations from the mean for the set of data 114, 142, 145, and 152.

Answer:

1. Value	Deviation
114	− 24.25
142	3.75
145	6.75
152	13.75

EXAMPLE 7

Find the deviations from the mean for the set of data 45, 63, 87, and 91.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{45 + 63 + 87 + 91}{4} = \frac{286}{4} = 71.5 \quad \text{Mean.}$$

To find the deviation from the mean, subtract the mean, $\bar{x}$, from each value of x. We arrange these values in a table.

	Values x_i	Deviations $x_i - \bar{x}$
	45	$45 - 71.5 =$ **−26.5**
	63	$63 - 71.5 =$ **−8.5**
	87	$87 - 71.5 =$ **15.5**
	91	$91 - 71.5 =$ **19.5**
Sum	286	0

See Exercises 26–34.

As we might expect, the sum of the deviations in the example equals zero because the sum of the negative deviations ($-26.5 + -8.5 = -35$) equals the sum of the positive deviations ($15.5 + 19.5 = 35$). To compensate for this situation, mathematicians employ a statistical measure called the **standard deviation,** which uses the square of each deviation from the mean. The square of a negative value is always positive. The sum of the squared deviations is divided by 1 less than the number of values, and the result is called the sample **variance:**

Standard deviation: a statistical measure of dispersion that shows how data vary about the mean; the square root of the variance

Variance: the sum of the squared deviations divided by 1 less than the number of values in the data set

$$\text{sample variance} = v = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}$$

The square root of the variance is the standard deviation. Various formulas exist for finding the standard deviation of a set of values, but we examine only one formula. Several calculations are necessary and are best organized in a table.

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

square root of the variance

To find the standard deviation of a set of sample data:

1. Find the mean, $\bar{x}$.
2. Find the deviation of each value from the mean: $(x_i - \bar{x})$
3. Square each deviation: $(x_i - \bar{x})^2$
4. Find the sum of the squared deviations: $\Sigma(x_i - \bar{x})^2$
5. Divide the sum of the squared deviations by *one less than* the number of values in the data set. The quotient is called the *sample variance:*

$$v = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}.$$

6. Find the sample standard deviation by taking the square root of the sample variance:

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}}.$$

EXAMPLE 8

Find the sample standard deviation to the nearest tenth for the values 45, 63, 87, and 91.

$\bar{x} =$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
45		
63		
87		
91		
Sum Σ		

$n =$

$$v = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}$$

Square root of the variance is the standard deviation.

21.6

Did You Know? The terms *small* and *large* are subjective terms that vary with data sets and industry standards.

A small standard deviation indicates that the mean is a typical value in the data set. A large standard deviation indicates that the mean is not typical, and other statistical measures should be examined to better understand the characteristics of the data set.

Let's examine the various statistics for the data set on a number line (Fig. 6–16).

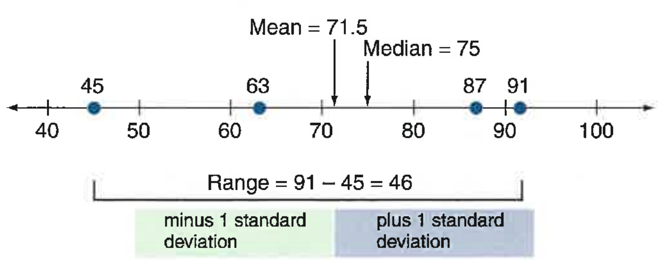


FIGURE 6–16

$$\text{Median} = \frac{63 + 87}{2} = \frac{150}{2} = 75.$$

Mean − 1 standard deviation $= 71.5 - 21.6 = 49.9.$

Mean + 1 standard deviation $= 71.5 + 21.6 = 93.1.$

We can confirm visually that the dispersion of the data is broad and the mean is not a typical value in the data set.

Normal distribution: an arrangement of a data set in which most of the values cluster around the mean and the rest taper off symmetrically toward the two ends. The graph representing a normal distribution is a bell-shaped curve

Another interpretation of the standard deviation is in its relationship to the normal distribution. A **normal distribution** is an arrangement of a data set in which most of the values cluster around the mean and the rest taper off symmetrically toward the two ends. The graph representing a normal distribution is a bell-shaped curve. If we assume that data are normally distributed, we make speculations about where data are located in relation to the population mean of the data set. Procedures for determining whether data are normally distributed are presented in advanced studies of statistics.

The graph of a normal distribution is a bell-shaped curve, as in Fig. 6–17. The curve is *symmetrical*; that is, if folded at the highest point of the curve, the two halves would match. The mean of the data set is at the highest point or fold line. Then, half the data (50%) are to the left or *below* the mean and half the data (50%) are to the right or *above* the mean. Other characteristics of the normal distribution are:

68.26% of the data are within 1 standard deviation of the mean.
95.44% of the data are within 2 standard deviations of the mean.
99.74% of the data are within 3 standard deviations of the mean.

Commonly referred to as 68-95-99.7

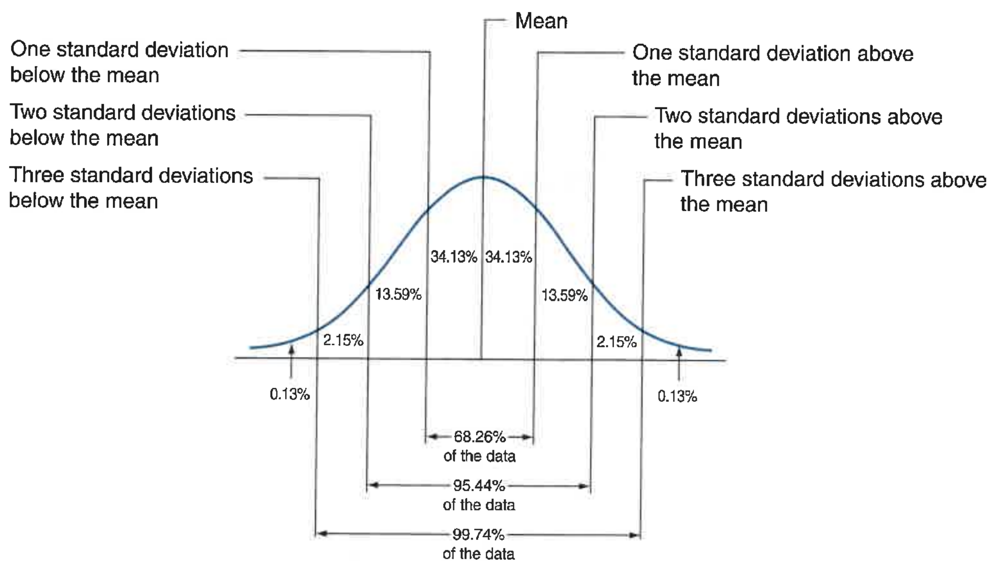


FIGURE 6–17

To solve applied problems involving the mean, standard deviation, and a normal distribution of a population:

1. Locate the mean and the desired values on the normal curve.

$$\frac{\text{value} - \text{population mean}}{\text{population standard deviation}}$$
$$= \text{number of standard deviations from (above or below) the mean}$$

2. Highlight the desired regions of the normal curve based on the conditions of the problem.
3. Add the percents associated with the highlighted regions in Step 2.

STOP AND CHECK

1. A brand of automobile tires has a population mean life of 50,000 miles with a standard deviation of 5,000 miles. In an inventory of 250 tires, what percent do you expect to last no more than 45,000 miles?

Answer:
1. 15.87% of the tires should last 45,000 miles or less.

Andrey Popov/Shutterstock

EXAMPLE 9

AUTO An Auto Zone Duralast Gold automobile battery has a population mean life of 46 months with a population standard deviation of 4 months. In an order of 100 batteries, what percent do you expect to last less than 54 months?

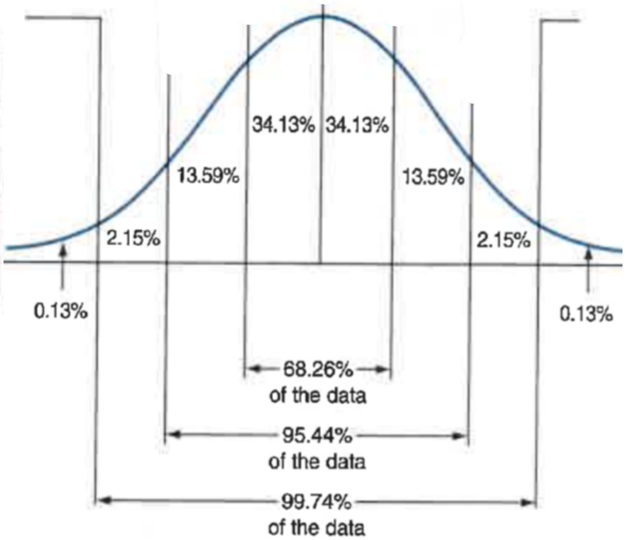
(54 months or more)

$$\frac{\text{value} - \text{population mean}}{\text{population standard deviation}}$$

= number of standard deviations *from* (above or below) the mean

Looking at the number of standard deviations from the mean gives you an idea about the answer.

46mo



2

50% of the data is above the mean and 50% below the mean.

STOP AND CHECK

1. Using the facts in the previous Stop and Check problem, how many of the 250 tires do you expect to last longer than 45,000 miles?

Answer:
1. 210 tires are expected to last 45,000 miles or more.

EXAMPLE 10

AUTO Using the facts from Example 9, how many batteries do you expect to last at least 54 months? Round to the nearest battery. Order is 100 batteries

2.28%

How many out of the order?

2

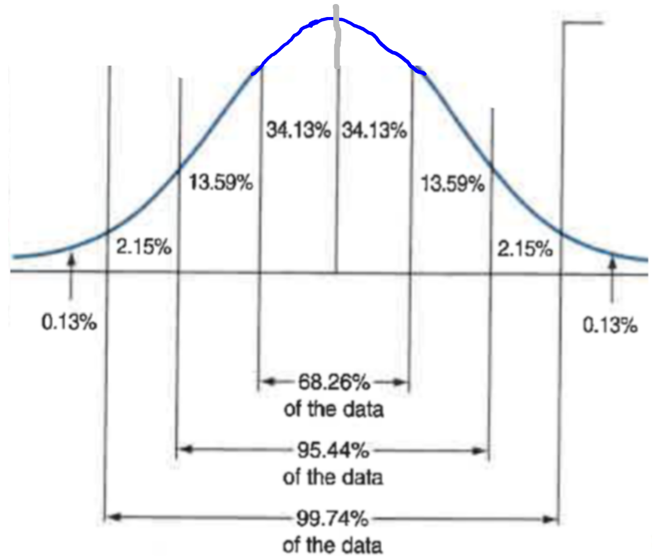
6.3.37

Question Help

Pediatricians work an average of 50 h per week. The standard deviation is 15 hours. What percentage of pediatricians work more than 80 h per week?

____% of pediatricians work more than 80 h per week.

____% of pediatricians work less than 35 h per week.



2.28
15.86

6-3 EXERCISES

MyLab Math For additional practice go to your study plan in MyLab Math.

1 Find the range for each data set. *See Example 1.*

1. 22, 36, 41, 41, 17
2. 28, 33, 36, 13, 28
3. 10, 23, 12, 17, 13, 16
4. 23, 23, 18, 32, 29, 14
5. \$25, \$15, \$25, \$40, \$19
6. \$36, \$44, \$26, \$52, \$19
7. 23°F, 37°F, 29°F, 54°F, 46°F, 71°F, 67°F

Table 6-8 Production in Tons of Sweet Cherries by State

State	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012*
CA	55,500	65,600	73,000	52,700	42,100	85,000	86,000	78,000	97,000	75,000	85,000
ID	1,700	2,900	3,100	1,700	3,800	1,500	1,900	6,000	1,900	2,800	4,000
MI	2,700	13,000	24,700	27,000	21,500	27,300	26,500	28,700	15,100	18,600	3,300
MT	2,220	2,060	2,360	1,230	2,400	2,440	1,560	23,900	2,470	2,015	
NY	350	600	900	800	960	1,190	1,050	1,240	1,000	700	250
OR	31,000	41,000	43,000	28,600	50,000	35,000	30,000	67,000	38,150	43,200	53,000
UT	400	2,200	1,600	1,800	1,800	1,250	50	1,540	1,100	800	1,600
WA	86,000	118,000	137,000	137,000	171,000	157,000	100,000	245,000	156,000	200,000	235,000
Total U.S.	180,225	245,700	282,060	250,830	293,560	310,680	247,060	429,870	312,720	343,115	382,150

*Forecast for 2012

Source: National Agricultural Statistics Service (NASS), Agricultural Statistics Board, U.S. Department of Agriculture (USDA), 2012.

Use Table 6–8 to answer Exercises 8–17. *See Example 1.*

8. **AG/H** Find the range for cherry production in states in 2003.

9. **AG/H** Find the range for cherry production in states in 2004.

10. **AG/H** Find the range for cherry production in states in 2011.

11. **AG/H** Find the range for cherry production for Washington State across the years 2002–2012.

12. **AG/H** Find the range for cherry production for the state of California across the years 2002–2012.

13. **AG/H** Find the range for cherry production for New York State across the years 2002–2012.

See Example 2.

14. Find the midrange for cherry production in states in 2002.

15. Find the midrange for cherry production in states in 2011.

16. Find the midrange for cherry production in the state of Washington (WA) across the years 2002–2012.

17. Find the midrange for cherry production in the state of Oregon (OR) across the years 2002–2012.

2 *See Example 3.*

18. Complete the table below to make a relative frequency distribution, a cumulative frequency distribution, and a cumulative relative frequency distribution for the data shown in the table.

Class	Class Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
30–39	4			
40–49	3			
50–59	2			
60–69	8			
70–79	5			
80–89	6			
90–99	4			

19. Complete the table below to make a relative frequency distribution, a cumulative frequency distribution, and a cumulative relative frequency distribution for the data shown in the table.

Class	Class Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
100–119	45			
120–139	54			
140–159	72			
160–179	63			
180–199	81			
200–219	45			

See Example 4.

20. Find the quartiles for the set of AlC measures (an AlC test measures average blood sugar levels over time by taking a sample of hemoglobin AlC cells) for a 56-year-old person.

6.2 5.9 7.1 6.1 5.8 4.9 6.3 6.4 6.7 5.8 5.9 4.5

21. Find the quartiles for the set of total blood serum cholesterol levels recorded by a lab.

205 194 188 220 245 218 255 231 214 199
187 201 238 229 231 267 188 191 203 205

See Example 5.

22. Find the inner-quartile range for the data in Exercise 20.

23. Find the inner-quartile range for the data in Exercise 21.

See Example 6.

24. Find the following percentiles for the scores in Exercise 20:
(a) 25% **(b)** 95%

25. Find the following percentiles for scores in Exercise 21:
(a) 40% **(b)** 98%

3 Find the standard deviation for each data set. Round to the nearest hundredth. *See Examples 7 and 8.*

26. 12, 14, 16, 18, 20

27. 68, 54, 73, 69

28. 32°F, 41°F, 54°F

29. \$27, \$32, \$65, \$29, \$21

30. Respiration rates: 16, 24, 20

31. Pulse rates: 68, 84, 76

Use Table 6–8 to answer Exercises 32–34.

32. **AG/H** Find the standard deviation for cherry production for Washington State for the years reported. Round to the nearest ton.

33. **AG/H** Find the standard deviation for cherry production for California for the years reported. Round to the nearest ton.

34. **AG/H** Find the standard deviation for cherry production for New York State for the years reported. Round to the nearest ton.

See Example 9.

35. **HLTH/N** The population mean length of a hospital stay for surgery is 5.8 days and the standard deviation is 1.9 days. What percentage of patients are hospitalized for 3.9 days or less?

36. **HLTH/N** Research has documented that the mean brain weight of people with Alzheimer's disease is 1,076.8 g and the standard deviation is 105.8. What percent of patients have a brain weight greater than 1,288.4 g?

37. **HLTH/N** Pediatricians work an average of 50 h per week. The standard deviation is 16 hours. What percentage of pediatricians work less than 18 h per week?

See Example 10.

38. **HLTH/N** In a sample of 80 pediatricians, how many work less than 18 h per week if the mean is 50 h per week and the standard deviation is 16 h?

39. **HLTH/N** In a sample of 200 surgery patients, how many are expected to stay 9.7 or more days if the mean length of stay is 5.8 days and the standard deviation is 3.9 days?

40. **HLTH/N** The mean brain weight of people with Alzheimer's disease is 1,076.8 g and the standard deviation is 105.8. In a sample of 500 Alzheimer's patients, how many will have a brain weight less than 865.2 g?