3–3 Measures of Variation

In statistics, to describe the data set accurately, statisticians must know more than the measures of central tendency. Consider Example 3–18.

Example 3–18

A testing lab wishes to test two experimental brands of outdoor paint to see how long each will last before fading. The testing lab makes 6 gallons of each paint to test. Since different chemical agents are added to each group and only six cans are involved, these two groups constitute two small populations. The results (in months) are shown. Find the mean of each group.

<table>
<thead>
<tr>
<th>Brand A</th>
<th>Brand B</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

Solution

The mean for brand A is

\[
\mu = \frac{\sum X}{N} = \frac{210}{6} = 35 \text{ months}
\]

The mean for brand B is

\[
\mu = \frac{\sum X}{N} = \frac{210}{6} = 35 \text{ months}
\]
Since the means are equal in Example 3–18, one might conclude that both brands of paint last equally well. However, when the data sets are examined graphically, a somewhat different conclusion might be drawn. See Figure 3–2.

As Figure 3–2 shows, even though the means are the same for both brands, the spread, or variation, is quite different. Figure 3–2 shows that brand B performs more consistently; it is less variable. For the spread or variability of a data set, three measures are commonly used: range, variance, and standard deviation. Each measure will be discussed in this section.

**Range**
The range is the simplest of the three measures and is defined now.

The **range** is the highest value minus the lowest value. The symbol $R$ is used for the range.

$$ R = \text{highest value} - \text{lowest value} $$

**Example 3–19** Find the ranges for the paints in Example 3–18.

**Solution**
For brand A, the range is

$$ R = 60 - 10 = 50 \text{ months} $$

For brand B, the range is

$$ R = 45 - 25 = 20 \text{ months} $$

Make sure the range is given as a single number.

The range for brand A shows that 50 months separate the largest data value from the smallest data value. For brand B, 20 months separate the largest data value from the smallest data value, which is less than one-half of brand A’s range.
One extremely high or one extremely low data value can affect the range markedly, as shown in Example 3–20.

**Example 3–20**

The salaries for the staff of the XYZ Manufacturing Co. are shown here. Find the range.

<table>
<thead>
<tr>
<th>Staff</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owner</td>
<td>$100,000</td>
</tr>
<tr>
<td>Manager</td>
<td>40,000</td>
</tr>
<tr>
<td>Sales representative</td>
<td>30,000</td>
</tr>
<tr>
<td>Workers</td>
<td>25,000</td>
</tr>
<tr>
<td></td>
<td>15,000</td>
</tr>
<tr>
<td></td>
<td>18,000</td>
</tr>
</tbody>
</table>

**Solution**

The range is \( R = 100,000 - 15,000 = 85,000 \). Since the owner’s salary is included in the data for Example 3–20, the range is a large number. To have a more meaningful statistic to measure the variability, statisticians use measures called the variance and standard deviation.

**Population Variance and Standard Deviation**

Before the variance and standard deviation are defined formally, the computational procedure will be shown, since the definition is derived from the procedure.

**Rounding Rule for the Standard Deviation**
The rounding rule for the standard deviation is the same as that for the mean. The final answer should be rounded to one more decimal place than that of the original data.

**Example 3–21**

Find the variance and standard deviation for the data set for brand A paint in Example 3–18.

10, 60, 50, 30, 40, 20

**Solution**

**Step 1** Find the mean for the data.

\[
\mu = \frac{\sum X}{N} = \frac{10 + 60 + 50 + 30 + 40 + 20}{6} = \frac{210}{6} = 35
\]

**Step 2** Subtract the mean from each data value.

\[
10 - 35 = -25 \quad 50 - 35 = +15 \quad 40 - 35 = +5
\]

\[
60 - 35 = +25 \quad 30 - 35 = -5 \quad 20 - 35 = -15
\]

**Step 3** Square each result.

\[
(-25)^2 = 625 \quad (+15)^2 = 225 \quad (+5)^2 = 25
\]

\[
(+25)^2 = 625 \quad (-5)^2 = 25 \quad (-15)^2 = 225
\]

**Step 4** Find the sum of the squares.

\[
625 + 625 + 225 + 25 + 25 + 225 = 1750
\]
Step 5  Divide the sum by \( N \) to get the variance.

Variance = \( 1750 \div 6 = 291.7 \)

Step 6  Take the square root of the variance to get the standard deviation. Hence, the standard deviation equals \( \sqrt{291.7} \), or 17.1. It is helpful to make a table.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values ((X))</td>
<td>(X - \mu)</td>
<td>((X - \mu)^2)</td>
</tr>
<tr>
<td>10</td>
<td>-25</td>
<td>625</td>
</tr>
<tr>
<td>60</td>
<td>+25</td>
<td>625</td>
</tr>
<tr>
<td>50</td>
<td>+15</td>
<td>225</td>
</tr>
<tr>
<td>30</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>40</td>
<td>+5</td>
<td>25</td>
</tr>
<tr>
<td>20</td>
<td>-15</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1750</td>
</tr>
</tbody>
</table>

Column A contains the raw data \( X \). Column B contains the differences \( X - \mu \) obtained in step 2. Column C contains the squares of the differences obtained in step 3.

The preceding computational procedure reveals several things. First, the square root of the variance gives the standard deviation; and vice versa, squaring the standard deviation gives the variance. Second, the variance is actually the average of the square of the distance that each value is from the mean. Therefore, if the values are near the mean, the variance will be small. In contrast, if the values are far from the mean, the variance will be large.

One might wonder why the squared distances are used instead of the actual distances. One reason is that the sum of the distances will always be zero. To verify this result for a specific case, add the values in column B of the table in Example 3–21. When each value is squared, the negative signs are eliminated.

Finally, why is it necessary to take the square root? The reason is that since the distances were squared, the units of the resultant numbers are the squares of the units of the original raw data. Finding the square root of the variance puts the standard deviation in the same units as the raw data.

When you are finding the square root, always use its positive or principal value, since the variance and standard deviation of a data set can never be negative.

The variance is the average of the squares of the distance each value is from the mean. The symbol for the population variance is \( \sigma^2 \) (\( \sigma \) is the Greek lowercase letter sigma). The formula for the population variance is

\[
\sigma^2 = \frac{\sum(X - \mu)^2}{N}
\]

where

- \( X \) = individual value
- \( \mu \) = population mean
- \( N \) = population size

The standard deviation is the square root of the variance. The symbol for the population standard deviation is \( \sigma \). The corresponding formula for the population standard deviation is

\[
\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(X - \mu)^2}{N}}
\]
Find the variance and standard deviation for brand B paint data in Example 3–18. The months were 35, 45, 30, 35, 40, 25

**Solution**

**Step 1** Find the mean.

\[
\mu = \frac{\sum X}{N} = \frac{35 + 45 + 30 + 35 + 40 + 25}{6} = \frac{210}{6} = 35
\]

**Step 2** Subtract the mean from each value, and place the result in column B of the table.

**Step 3** Square each result and place the squares in column C of the table.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>B</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>-5</td>
<td>25</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>25</td>
<td>-10</td>
<td>100</td>
</tr>
</tbody>
</table>

**Step 4** Find the sum of the squares in column C.

\[
\Sigma (X - \mu)^2 = 0 + 100 + 25 + 0 + 25 + 100 = 250
\]

**Step 5** Divide the sum by \(N\) to get the variance.

\[
\sigma^2 = \frac{\Sigma (X - \mu)^2}{N} = \frac{250}{6} = 41.7
\]

**Step 6** Take the square root to get the standard deviation.

\[
\sigma = \sqrt{\frac{\Sigma (X - \mu)^2}{N}} = \sqrt{41.7} = 6.5
\]

Hence, the standard deviation is 6.5.

Since the standard deviation of brand A is 17.1 (see Example 3–21) and the standard deviation of brand B is 6.5, the data are more variable for brand A. In summary, when the means are equal, the larger the variance or standard deviation is, the more variable the data are.

**Sample Variance and Standard Deviation**

When computing the variance for a sample, one might expect the following expression to be used:

\[
\frac{\Sigma (X - \bar{X})^2}{n}
\]

where \(\bar{X}\) is the sample mean and \(n\) is the sample size. This formula is not usually used, however, since in most cases the purpose of calculating the statistic is to estimate the...
corresponding parameter. For example, the sample mean $\bar{X}$ is used to estimate the population mean $\mu$. The expression

$$\frac{\sum (X - \bar{X})^2}{n}$$

do not give the best estimate of the population variance because when the population is large and the sample is small (usually less than 30), the variance computed by this formula usually underestimates the population variance. Therefore, instead of dividing by $n$, find the variance of the sample by dividing by $n - 1$, giving a slightly larger value and an unbiased estimate of the population variance.

The formula for the sample variance, denoted by $s^2$, is

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$$

where

$\bar{X} = \text{sample mean}$

$n = \text{sample size}$

To find the standard deviation of a sample, one must take the square root of the sample variance, which was found by using the preceding formula.

**Formula for the Sample Standard Deviation**

The standard deviation of a sample (denoted by $s$) is

$$s = \sqrt{s^2} = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

where

$X = \text{individual value}$

$\bar{X} = \text{sample mean}$

$n = \text{sample size}$

Shortcut formulas for computing the variance and standard deviation are presented next and will be used in the remainder of the chapter and in the exercises. These formulas are mathematically equivalent to the preceding formulas and do not involve using the mean. They save time when repeated subtracting and squaring occur in the original formulas. They are also more accurate when the mean has been rounded.

**Shortcut or Computational Formulas for $s^2$ and $s$**

The shortcut formulas for computing the variance and standard deviation for data obtained from samples are as follows.

**Variance**

$$s^2 = \frac{\sum X^2 - [(\sum X)^2/n]}{n - 1}$$

**Standard deviation**

$$s = \sqrt{\frac{\sum X^2 - [(\sum X)^2/n]}{n - 1}}$$
Examples 3–23 and 3–24 explain how to use the shortcut formulas.

**Example 3–23**

Find the sample variance and standard deviation for the amount of European auto sales for a sample of 6 years shown. The data are in millions of dollars.

11.2, 11.9, 12.0, 12.8, 13.4, 14.3

*Source: USA TODAY.*

**Solution**

**Step 1** Find the sum of the values.

\[ \sum X = 11.2 + 11.9 + 12.0 + 12.8 + 13.4 + 14.3 = 75.6 \]

**Step 2** Square each value and find the sum.

\[ \sum X^2 = 11.2^2 + 11.9^2 + 12.0^2 + 12.8^2 + 13.4^2 + 14.3^2 = 958.94 \]

**Step 3** Substitute in the formulas and solve.

\[ s^2 = \frac{\sum X^2 - \left(\frac{\sum X}{n}\right)^2}{n - 1} = \frac{958.94 - \left(\frac{75.6}{6}\right)^2}{5} = 1.28 \]

The variance of the sample is 1.28.

\[ s = \sqrt{1.28} = 1.13 \]

Hence, the sample standard deviation is 1.13.

*Note that \( \sum X^2 \) is not the same as \((\sum X)^2\). The notation \( \sum X^2 \) means to square the values first, then sum; \((\sum X)^2\) means to sum the values first, then square the sum.*

**Variance and Standard Deviation for Grouped Data**

The procedure for finding the variance and standard deviation for grouped data is similar to that for finding the mean for grouped data, and it uses the midpoints of each class.

**Example 3–24**

Find the variance and the standard deviation for the frequency distribution of the data in Example 2–7. The data represent the number of miles that 20 runners ran during one week.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5–10.5</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>10.5–15.5</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>15.5–20.5</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>20.5–25.5</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>25.5–30.5</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>30.5–35.5</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>35.5–40.5</td>
<td>2</td>
<td>38</td>
</tr>
</tbody>
</table>
**Solution**

**Step 1** Make a table as shown, and find the midpoint of each class.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency (f)</th>
<th>Midpoint ($X_m$)</th>
<th>$f \cdot X_m$</th>
<th>$f \cdot X^2_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5–10.5</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>10.5–15.5</td>
<td>2</td>
<td>13</td>
<td>26</td>
<td>338</td>
</tr>
<tr>
<td>15.5–20.5</td>
<td>3</td>
<td>18</td>
<td>54</td>
<td>972</td>
</tr>
<tr>
<td>20.5–25.5</td>
<td>5</td>
<td>23</td>
<td>115</td>
<td>2,645</td>
</tr>
<tr>
<td>25.5–30.5</td>
<td>4</td>
<td>28</td>
<td>112</td>
<td>3,136</td>
</tr>
<tr>
<td>30.5–35.5</td>
<td>3</td>
<td>33</td>
<td>99</td>
<td>3,267</td>
</tr>
<tr>
<td>35.5–40.5</td>
<td>2</td>
<td>38</td>
<td>76</td>
<td>2,888</td>
</tr>
</tbody>
</table>

**Step 2** Multiply the frequency by the midpoint for each class, and place the products in column D.

$1 \cdot 8 = 8 \quad 2 \cdot 13 = 26 \quad \ldots \quad 2 \cdot 38 = 76$

**Step 3** Multiply the frequency by the square of the midpoint, and place the products in column E.

$1 \cdot 8^2 = 64 \quad 2 \cdot 13^2 = 338 \quad \ldots \quad 2 \cdot 38^2 = 2888$

**Step 4** Find the sums of columns B, D, and E. The sum of column B is $n$, the sum of column D is $\Sigma f \cdot X_m$, and the sum of column E is $\Sigma f \cdot X^2_m$. The completed table is shown.

<table>
<thead>
<tr>
<th>A Class</th>
<th>B Frequency</th>
<th>C Midpoint</th>
<th>D $f \cdot X_m$</th>
<th>E $f \cdot X^2_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5–10.5</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>10.5–15.5</td>
<td>2</td>
<td>13</td>
<td>26</td>
<td>338</td>
</tr>
<tr>
<td>15.5–20.5</td>
<td>3</td>
<td>18</td>
<td>54</td>
<td>972</td>
</tr>
<tr>
<td>20.5–25.5</td>
<td>5</td>
<td>23</td>
<td>115</td>
<td>2,645</td>
</tr>
<tr>
<td>25.5–30.5</td>
<td>4</td>
<td>28</td>
<td>112</td>
<td>3,136</td>
</tr>
<tr>
<td>30.5–35.5</td>
<td>3</td>
<td>33</td>
<td>99</td>
<td>3,267</td>
</tr>
<tr>
<td>35.5–40.5</td>
<td>2</td>
<td>38</td>
<td>76</td>
<td>2,888</td>
</tr>
</tbody>
</table>

$n = 20 \quad \Sigma f \cdot X_m = 490 \quad \Sigma f \cdot X^2_m = 13,310$

**Step 5** Substitute in the formula and solve for $s^2$ to get the variance.

$$s^2 = \frac{\Sigma f \cdot X^2_m - [(\Sigma f \cdot X_m)^2/n]}{n - 1}$$

$$= \frac{13,310 - [490]^2/20}{20 - 1} = 68.7$$

**Step 6** Take the square root to get the standard deviation.

$$s = \sqrt{68.7} = 8.3$$

Be sure to use the number found in the sum of column B (i.e., the sum of the frequencies) for $n$. Do not use the number of classes.

The steps for finding the variance and standard deviation for grouped data are summarized in this Procedure Table.
The three measures of variation are summarized in Table 3–2.

### Procedure Table

**Finding the Sample Variance and Standard Deviation for Grouped Data**

**Step 1** Make a table as shown, and find the midpoint of each class.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Frequency</td>
<td>Midpoint</td>
<td>$f \cdot X_m$</td>
<td>$f \cdot X_m^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Step 2** Multiply the frequency by the midpoint for each class, and place the products in column D.

**Step 3** Multiply the frequency by the square of the midpoint, and place the products in column E.

**Step 4** Find the sums of columns B, D, and E. (The sum of column B is $n$. The sum of column D is $\sum f \cdot X_m$. The sum of column E is $\sum f \cdot X_m^2$.)

**Step 5** Substitute in the formula and solve to get the variance.

$$s^2 = \frac{\sum f \cdot X_m^2 - (\sum f \cdot X_m)^2/n}{n - 1}$$

**Step 6** Take the square root to get the standard deviation.

### Unusual Stat

The average number of times that a man cries in a month is 1.4.

### Table 3–2 Summary of Measures of Variation

<table>
<thead>
<tr>
<th>Measure</th>
<th>Definition</th>
<th>Symbol(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>Distance between highest value and lowest value</td>
<td>$R$</td>
</tr>
<tr>
<td>Variance</td>
<td>Average of the squares of the distance that each value is from the mean</td>
<td>$\sigma^2$, $s^2$</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>Square root of the variance</td>
<td>$\sigma$, $s$</td>
</tr>
</tbody>
</table>

### Uses of the Variance and Standard Deviation

1. As previously stated, variances and standard deviations can be used to determine the spread of the data. If the variance or standard deviation is large, the data are more dispersed. This information is useful in comparing two (or more) data sets to determine which is more (most) variable.

2. The measures of variance and standard deviation are used to determine the consistency of a variable. For example, in the manufacture of fittings, such as nuts and bolts, the variation in the diameters must be small, or the parts will not fit together.

3. The variance and standard deviation are used to determine the number of data values that fall within a specified interval in a distribution. For example, Chebyshev’s theorem (explained later) shows that, for any distribution, at least 75% of the data values will fall within 2 standard deviations of the mean.

4. Finally, the variance and standard deviation are used quite often in inferential statistics. These uses will be shown in later chapters of this textbook.
Coefficient of Variation

Whenever two samples have the same units of measure, the variance and standard deviation for each can be compared directly. For example, suppose an automobile dealer wanted to compare the standard deviation of miles driven for the cars she received as trade-ins on new cars. She found that for a specific year, the standard deviation for Buicks was 422 miles and the standard deviation for Cadillacs was 350 miles. She could say that the variation in mileage was greater in the Buicks. But what if a manager wanted to compare the standard deviations of two different variables, such as the number of sales per salesperson over a 3-month period and the commissions made by these salespeople?

A statistic that allows one to compare standard deviations when the units are different, as in this example, is called the \textit{coefficient of variation}.

The \textbf{coefficient of variation}, denoted by \( CVar \), is the standard deviation divided by the mean. The result is expressed as a percentage.

\[
\text{For samples,} \quad CVar = \frac{s}{X} \cdot 100\% \\
\text{For populations,} \quad CVar = \frac{\sigma}{\mu} \cdot 100\%
\]

\begin{example}

\textbf{Example 3–25}

The mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is $5225, and the standard deviation is $773. Compare the variations of the two.

\textbf{Solution}

The coefficients of variation are

\[
CVar = \frac{s}{X} = \frac{5}{87} \cdot 100\% = 5.7\% \quad \text{sales}
\]

\[
CVar = \frac{773}{5225} \cdot 100\% = 14.8\% \quad \text{commissions}
\]

Since the coefficient of variation is larger for commissions, the commissions are more variable than the sales.

\end{example}

\begin{example}

\textbf{Example 3–26}

The mean for the number of pages of a sample of women’s fitness magazines is 132, with a variance of 23; the mean for the number of advertisements of a sample of women’s fitness magazines is 182, with a variance of 62. Compare the variations.

\textbf{Solution}

The coefficients of variation are

\[
CVar = \frac{\sqrt{23}}{132} \cdot 100\% = 3.6\% \quad \text{pages}
\]

\[
CVar = \frac{\sqrt{62}}{182} \cdot 100\% = 4.3\% \quad \text{advertisements}
\]

The number of advertisements is more variable than the number of pages since the coefficient of variation is larger for advertisements.

\end{example}
Range Rule of Thumb

The range can be used to approximate the standard deviation. The approximation is called the range rule of thumb.

The Range Rule of Thumb

A rough estimate of the standard deviation is

\[ s \approx \frac{\text{range}}{4} \]

In other words, if the range is divided by 4, an approximate value for the standard deviation is obtained. For example, the standard deviation for the data set 5, 8, 8, 9, 10, 12, and 13 is 2.7, and the range is 13 − 5 = 8. The range rule of thumb is \( s = 2 \). The range rule of thumb in this case underestimates the standard deviation somewhat; however, it is in the ballpark.

A note of caution should be mentioned here. The range rule of thumb is only an approximation and should be used when the distribution of data values is unimodal and roughly symmetric.

The range rule of thumb can be used to estimate the largest and smallest data values of a data set. The smallest data value will be approximately 2 standard deviations below the mean, and the largest data value will be approximately 2 standard deviations above the mean of the data set. The mean for the previous data set is 9.3; hence,

\[
\text{Smallest data value} = \bar{X} - 2s = 9.3 - 2(2.8) = 3.7 \\
\text{Largest data value} = \bar{X} + 2s = 9.3 + 2(2.8) = 14.9
\]

Notice that the smallest data value was 5, and the largest data value was 13. Again, these are rough approximations. For many data sets, almost all data values will fall within 2 standard deviations of the mean. Better approximations can be obtained by using Chebyshev’s theorem and the empirical rule. These are explained next.

Chebyshev’s Theorem

As stated previously, the variance and standard deviation of a variable can be used to determine the spread, or dispersion, of a variable. That is, the larger the variance or standard deviation, the more the data values are dispersed. For example, if two variables measured in the same units have the same mean, say, 70, and variable 1 has a standard deviation of 1.5 while variable 2 has a standard deviation of 10, then the data for variable 2 will be more spread out than the data for variable 1. Chebyshev’s theorem, developed by the Russian mathematician Chebyshev (1821–1894), specifies the proportions of the spread in terms of the standard deviation.

**Chebyshev’s theorem**  The proportion of values from a data set that will fall within \( k \) standard deviations of the mean will be at least \( 1 - \frac{1}{k^2} \), where \( k \) is a number greater than 1 (\( k \) is not necessarily an integer).

This theorem states that at least three-fourths, or 75%, of the data values will fall within 2 standard deviations of the mean of the data set. This result is found by substituting \( k = 2 \) in the expression.

\[ 1 - \frac{1}{k^2} \quad \text{or} \quad 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 75\% \]
For the example in which variable 1 has a mean of 70 and a standard deviation of 1.5, at least three-fourths, or 75%, of the data values fall between 67 and 73. These values are found by adding 2 standard deviations to the mean and subtracting 2 standard deviations from the mean, as shown:

\[ 70 + 2(1.5) = 70 + 3 = 73 \]

and

\[ 70 - 2(1.5) = 70 - 3 = 67 \]

For variable 2, at least three-fourths, or 75%, of the data values fall between 50 and 90. Again, these values are found by adding and subtracting, respectively, 2 standard deviations to and from the mean.

\[ 70 + 2(10) = 70 + 20 = 90 \]

and

\[ 70 - 2(10) = 70 - 20 = 50 \]

Furthermore, the theorem states that at least eight-ninths, or 88.89%, of the data values will fall within 3 standard deviations of the mean. This result is found by letting \( k = 3 \) and substituting in the expression.

\[ 1 - \frac{1}{k^2} \quad \text{or} \quad 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 88.89\% \]

For variable 1, at least eight-ninths, or 88.89%, of the data values fall between 65.5 and 74.5, since

\[ 70 + 3(1.5) = 70 + 4.5 = 74.5 \]

and

\[ 70 - 3(1.5) = 70 - 4.5 = 65.5 \]

For variable 2, at least eight-ninths, or 88.89%, of the data values fall between 40 and 100.

This theorem can be applied to any distribution regardless of its shape (see Figure 3–3).

Examples 3–27 and 3–28 illustrate the application of Chebyshev’s theorem.
The mean price of houses in a certain neighborhood is $50,000, and the standard deviation is $10,000. Find the price range for which at least 75% of the houses will sell.

**Solution**
Chebyshev’s theorem states that three-fourths, or 75%, of the data values will fall within 2 standard deviations of the mean. Thus,

\[
50,000 + 2(10,000) = 50,000 + 20,000 = 70,000
\]

and

\[
50,000 - 2(10,000) = 50,000 - 20,000 = 30,000
\]

Hence, at least 75% of all homes sold in the area will have a price range from $30,000 to $70,000.

Chebyshev’s theorem can be used to find the minimum percentage of data values that will fall between any two given values. The procedure is shown in Example 3–28.

A survey of local companies found that the mean amount of travel allowance for executives was $0.25 per mile. The standard deviation was $0.02. Using Chebyshev’s theorem, find the minimum percentage of the data values that will fall between $0.20 and $0.30.

**Solution**

**Step 1** Subtract the mean from the larger value.

\[
0.30 - 0.25 = 0.05
\]

**Step 2** Divide the difference by the standard deviation to get \( k \).

\[
k = \frac{0.05}{0.02} = 2.5
\]

**Step 3** Use Chebyshev’s theorem to find the percentage.

\[
1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} = 1 - \frac{1}{6.25} = 1 - 0.16 = 0.84 \quad \text{or} \quad 84\%
\]

Hence, at least 84% of the data values will fall between $0.20 and $0.30.

**The Empirical (Normal) Rule**
Chebyshev’s theorem applies to any distribution regardless of its shape. However, when a distribution is *bell-shaped* (or what is called *normal*), the following statements, which make up the empirical rule, are true.

- Approximately 68% of the data values will fall within 1 standard deviation of the mean.
- Approximately 95% of the data values will fall within 2 standard deviations of the mean.
- Approximately 99.7% of the data values will fall within 3 standard deviations of the mean.
For example, suppose that the scores on a national achievement exam have a mean of 480 and a standard deviation of 90. If these scores are normally distributed, then approximately 68% will fall between 390 and 570 \((480 - 90, 480 + 90)\). Approximately 95% of the scores will fall between 300 and 660 \((480 - 2 \times 90, 480 + 2 \times 90)\). Approximately 99.7% will fall between 210 and 750 \((480 - 3 \times 90, 480 + 3 \times 90)\). See Figure 3–4. (The empirical rule is explained in greater detail in Chapter 7.)

### Applying the Concepts 3–3

#### Blood Pressure

The table lists means and standard deviations. The mean is the number before the plus/minus, and the standard deviation is the number after the plus/minus. The results are from a study attempting to find the average blood pressure of older adults. Use the results to answer the questions.

<table>
<thead>
<tr>
<th>Normotensive</th>
<th>Hypertensive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong> (n = 1200)</td>
<td><strong>Women</strong> (n = 1400)</td>
</tr>
<tr>
<td>Age</td>
<td>55 ± 10</td>
</tr>
<tr>
<td>Blood pressure (mm Hg)</td>
<td></td>
</tr>
<tr>
<td>Systolic</td>
<td>123 ± 9</td>
</tr>
<tr>
<td>Diastolic</td>
<td>78 ± 7</td>
</tr>
</tbody>
</table>

1. Apply Chebyshev’s theorem to the systolic blood pressure of normotensive men. At least how many of the men in the study fall within 1 standard deviation of the mean?
2. At least how many of those men in the study fall within 2 standard deviations of the mean?

Assume that blood pressure is normally distributed among older adults. Answer the following questions, using the empirical rule instead of Chebyshev’s theorem.

3. Give ranges for the diastolic blood pressure (normotensive and hypertensive) of older women.
4. Do the normotensive, male, systolic blood pressure ranges overlap with the hypertensive, male, systolic, blood pressure ranges?

See page 170 for the answers.
1. What is the relationship between the variance and the standard deviation?

2. Why might the range not be the best estimate of variability?

3. What are the symbols used to represent the population variance and standard deviation?

4. What are the symbols used to represent the sample variance and standard deviation?

5. Why is the unbiased estimator of variance used?

6. The three data sets have the same mean and range, but is the variation the same? Prove your answer by computing the standard deviation. Assume the data were obtained from samples.
   a. 5, 7, 9, 11, 13, 15, 17
   b. 5, 6, 7, 11, 15, 16, 17
   c. 5, 5, 5, 11, 17, 17

For Exercises 7–13, find the range, variance, and standard deviation. Assume the data represent samples, and use the shortcut formula for the unbiased estimator to compute the variance and standard deviation.

7. The number of incidents where police were needed for a sample of 10 schools in Allegheny County is 7, 37, 3, 8, 48, 11, 6, 0, 10, 3. Are the data consistent or do they vary? Explain your answer.
   Source: U.S. Department of Education.

8. The increases (in cents) in cigarette taxes for 17 states in a 6-month period are
   60, 20, 40, 40, 45, 12, 34, 51, 30, 70, 42, 31, 69, 32, 8, 18, 50
   Use the range rule of thumb to estimate the standard deviation. Compare the estimate to the actual standard deviation.
   Source: Federation of Tax Administrators.

9. The normal daily high temperatures (in degrees Fahrenheit) in January for 10 selected cities are as follows.
   50, 37, 29, 54, 30, 61, 47, 38, 34, 61
   The normal monthly precipitation (in inches) for these same 10 cities is listed here.
   4.8, 2.6, 1.5, 1.8, 1.8, 3.3, 5.1, 1.1, 1.8, 2.5
   Which set is more variable?
   Source: *N.Y. Times Almanac*.

10. The total surface area (in square miles) for each of six selected Eastern states is listed here.
    28,995 PA    37,534 FL
    31,361 NY    27,087 VA
    20,966 ME    37,741 GA
    The total surface area for each of six selected Western states is listed (in square miles).
    72,964 AZ    70,763 NV
    101,510 CA   62,161 OR
    66,625 CO    54,339 UT
    Which set is more variable?
    Source: *N.Y. Times Almanac*.

11. Shown here are the numbers of stories in the 11 tallest buildings in St. Paul, Minnesota.
    32, 36, 46, 20, 32, 18, 16, 34, 26, 27, 26
    Shown here are the numbers of stories in the 11 tallest buildings in Chicago, Illinois.
    100, 100, 83, 60, 64, 65, 66, 74, 60, 67, 57
    Which data set is more variable?
    Source: *The World Almanac and Book of Facts*.

12. The following data are the prices of 1 gallon of premium gasoline in U.S. dollars in seven foreign countries.
    3.80, 3.80, 3.20, 3.57, 3.62, 3.74, 3.69
    Do you think the standard deviation of these data is representative of the population standard deviation of gasoline prices in all foreign countries? Explain your answer.
    Source: *Pittsburgh Post Gazette*.

13. The number of weeks on *The New York Times Best Sellers* list for hardcover fiction is
    1, 4, 2, 2, 3, 18, 5, 5, 10, 4, 3, 6, 2, 2, 22
    Use the range rule of thumb to estimate the standard deviation. Compare the estimate to the actual standard deviation.

14. Find the range, variance, and standard deviation for the distances of the home runs for McGwire and Sosa, using the data in Exercise 18 in Section 2–2. Compare the ranges and standard deviations. Decide which is more variable or if the variability is about the same. (Use individual data.)

15. Find the range, variance, and standard deviation for each data set in Exercise 11 of Section 3–2. Based on the results, which data set is more variable?
16. The Federal Highway Administration reported the number of deficient bridges in each state. Find the range, variance, and standard deviation.

<table>
<thead>
<tr>
<th>State</th>
<th>Bridges</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15,458</td>
</tr>
<tr>
<td>B</td>
<td>1,055</td>
</tr>
<tr>
<td>C</td>
<td>5,008</td>
</tr>
<tr>
<td>D</td>
<td>3,598</td>
</tr>
<tr>
<td>E</td>
<td>8,984</td>
</tr>
<tr>
<td>F</td>
<td>1,337</td>
</tr>
<tr>
<td>G</td>
<td>4,132</td>
</tr>
<tr>
<td>H</td>
<td>7,768</td>
</tr>
<tr>
<td>I</td>
<td>4,131</td>
</tr>
<tr>
<td>J</td>
<td>810</td>
</tr>
<tr>
<td>K</td>
<td>13,350</td>
</tr>
<tr>
<td>L</td>
<td>10,902</td>
</tr>
<tr>
<td>M</td>
<td>14,318</td>
</tr>
<tr>
<td>N</td>
<td>1,208</td>
</tr>
<tr>
<td>O</td>
<td>97–159</td>
</tr>
<tr>
<td>P</td>
<td>34–96</td>
</tr>
<tr>
<td>Q</td>
<td>160–222</td>
</tr>
<tr>
<td>R</td>
<td>223–285</td>
</tr>
<tr>
<td>S</td>
<td>286–348</td>
</tr>
<tr>
<td>T</td>
<td>349–411</td>
</tr>
<tr>
<td>U</td>
<td>412–474</td>
</tr>
<tr>
<td>V</td>
<td>475–537</td>
</tr>
<tr>
<td>W</td>
<td>538–600</td>
</tr>
</tbody>
</table>

Source: USA TODAY.

17. Find the range, variance, and standard deviation for the data in Exercise 17 of Section 2–2.

For Exercises 18 through 27, find the variance and standard deviation.

18. For 108 randomly selected college students, this exam score frequency distribution was obtained.

<table>
<thead>
<tr>
<th>Class limits</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>90–98</td>
<td>6</td>
</tr>
<tr>
<td>99–107</td>
<td>22</td>
</tr>
<tr>
<td>108–116</td>
<td>43</td>
</tr>
<tr>
<td>117–125</td>
<td>28</td>
</tr>
<tr>
<td>126–134</td>
<td>9</td>
</tr>
</tbody>
</table>

19. The costs per load (in cents) of 35 laundry detergents tested by a consumer organization are shown here.

<table>
<thead>
<tr>
<th>Class limits</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>13–19</td>
<td>2</td>
</tr>
<tr>
<td>20–26</td>
<td>7</td>
</tr>
<tr>
<td>27–33</td>
<td>12</td>
</tr>
<tr>
<td>34–40</td>
<td>5</td>
</tr>
<tr>
<td>41–47</td>
<td>6</td>
</tr>
<tr>
<td>48–54</td>
<td>1</td>
</tr>
<tr>
<td>55–61</td>
<td>0</td>
</tr>
<tr>
<td>62–68</td>
<td>2</td>
</tr>
</tbody>
</table>

20. Thirty automobiles were tested for fuel efficiency (in miles per gallon). This frequency distribution was obtained.

<table>
<thead>
<tr>
<th>Class boundaries</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5–12.5</td>
<td>3</td>
</tr>
<tr>
<td>12.5–17.5</td>
<td>5</td>
</tr>
<tr>
<td>17.5–22.5</td>
<td>15</td>
</tr>
<tr>
<td>22.5–27.5</td>
<td>5</td>
</tr>
<tr>
<td>27.5–32.5</td>
<td>2</td>
</tr>
</tbody>
</table>

21. The data show the number of murders in 25 selected cities.

<table>
<thead>
<tr>
<th>Class limits</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>34–96</td>
<td>13</td>
</tr>
<tr>
<td>97–159</td>
<td>2</td>
</tr>
<tr>
<td>160–222</td>
<td>0</td>
</tr>
<tr>
<td>223–285</td>
<td>5</td>
</tr>
<tr>
<td>286–348</td>
<td>1</td>
</tr>
<tr>
<td>349–411</td>
<td>1</td>
</tr>
<tr>
<td>412–474</td>
<td>0</td>
</tr>
<tr>
<td>475–537</td>
<td>1</td>
</tr>
<tr>
<td>538–600</td>
<td>2</td>
</tr>
</tbody>
</table>

22. In a study of reaction times to a specific stimulus, a psychologist recorded these data (in seconds).

<table>
<thead>
<tr>
<th>Class limits</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1–2.7</td>
<td>12</td>
</tr>
<tr>
<td>2.8–3.4</td>
<td>13</td>
</tr>
<tr>
<td>3.5–4.1</td>
<td>7</td>
</tr>
<tr>
<td>4.2–4.8</td>
<td>5</td>
</tr>
<tr>
<td>4.9–5.5</td>
<td>2</td>
</tr>
<tr>
<td>5.6–6.2</td>
<td>1</td>
</tr>
</tbody>
</table>

23. Eighty randomly selected lightbulbs were tested to determine their lifetimes (in hours). This frequency distribution was obtained.

<table>
<thead>
<tr>
<th>Class boundaries</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>52.5–63.5</td>
<td>6</td>
</tr>
<tr>
<td>63.5–74.5</td>
<td>12</td>
</tr>
<tr>
<td>74.5–85.5</td>
<td>25</td>
</tr>
<tr>
<td>85.5–96.5</td>
<td>18</td>
</tr>
<tr>
<td>96.5–107.5</td>
<td>14</td>
</tr>
<tr>
<td>107.5–118.5</td>
<td>5</td>
</tr>
</tbody>
</table>

24. The data represent the murder rate per 100,000 individuals in a sample of selected cities in the United States.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>5–11</td>
<td>8</td>
</tr>
<tr>
<td>12–18</td>
<td>5</td>
</tr>
<tr>
<td>19–25</td>
<td>7</td>
</tr>
<tr>
<td>26–32</td>
<td>1</td>
</tr>
<tr>
<td>33–39</td>
<td>1</td>
</tr>
<tr>
<td>40–46</td>
<td>3</td>
</tr>
</tbody>
</table>

Source: FBI and U.S. Census Bureau.

25. Eighty randomly selected batteries were tested to determine their lifetimes (in hours). The following frequency distribution was obtained.
Can it be concluded that the lifetimes of these brands of batteries are consistent?

26. Find the variance and standard deviation for the two distributions in Exercise 8 in Section 2–3 and Exercise 18 in Section 2–3. Compare the variation of the data sets. Decide if one data set is more variable than the other.

27. This frequency distribution represents the data obtained from a sample of word processor repairers. The values are the days between service calls on 80 machines.

<table>
<thead>
<tr>
<th>Class boundaries</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.5–28.5</td>
<td>5</td>
</tr>
<tr>
<td>28.5–31.5</td>
<td>9</td>
</tr>
<tr>
<td>31.5–34.5</td>
<td>32</td>
</tr>
<tr>
<td>34.5–37.5</td>
<td>20</td>
</tr>
<tr>
<td>37.5–40.5</td>
<td>12</td>
</tr>
<tr>
<td>40.5–43.5</td>
<td>2</td>
</tr>
</tbody>
</table>

28. The average score of the students in one calculus class is 110, with a standard deviation of 5; the average score of students in a statistics class is 106, with a standard deviation of 4. Which class is more variable in terms of scores?

29. The data show the lengths (in feet) of suspension bridges in the eastern part of North America and the western part of North America. Compare the variability of the two samples, using the coefficient of variation.

<table>
<thead>
<tr>
<th>East</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4260</td>
<td>3500</td>
<td>2300</td>
</tr>
<tr>
<td></td>
<td>2150</td>
<td>2000</td>
<td>1750</td>
</tr>
<tr>
<td>West</td>
<td>4200</td>
<td>2800</td>
<td>2310</td>
</tr>
<tr>
<td></td>
<td>1550</td>
<td>1500</td>
<td>1207</td>
</tr>
</tbody>
</table>


30. The average score on an English final examination was 85, with a standard deviation of 5; the average score on a history final exam was 110, with a standard deviation of 8. Which class was more variable?

31. The average age of the accountants at Three Rivers Corp. is 26 years, with a standard deviation of 6 years; the average salary of the accountants is $31,000, with a standard deviation of $4000. Compare the variations of age and income.

32. Using Chebyshev’s theorem, solve these problems for a distribution with a mean of 80 and a standard deviation of 10.

33. The mean of a distribution is 20 and the standard deviation is 2. Use Chebyshev’s theorem.

a. At least what percentage of the values will fall between 18 and 22?

b. At least what percentage of the values will fall between 14 and 26?

34. In a distribution of 200 values, the mean is 50 and the standard deviation is 5. Use Chebyshev’s theorem.

a. At least how many values will fall between 40 and 60?

b. At most how many values will be less than 35 or more than 65?

35. A sample of the hourly wages of employees who work in restaurants in a large city has a mean of $5.02 and a standard deviation of $0.09. Using Chebyshev’s theorem, find the range in which at least 50% of the data values will fall.

36. A sample of the labor costs per hour to assemble a certain product has a mean of $2.60 and a standard deviation of $0.15. Using Chebyshev’s theorem, find the range in which at least 88.89% of the data will lie.

37. A survey of a number of the leading brands of cereal shows that the mean content of potassium per serving is 95 milligrams, and the standard deviation is 2 milligrams. Find the range in which at least 75% of the data will fall. Use Chebyshev’s theorem.

38. The average score on a special test of knowledge of wood refinishing has a mean of 53 and a standard deviation of 6. Using Chebyshev’s theorem, find the range in which at least 75% of the scores will lie.

39. The average of the number of trials it took a sample of mice to learn to traverse a maze was 12. The standard deviation was 3. Using Chebyshev’s theorem, find the minimum percentage of data values that will fall in the range of 4 to 20 trials.

40. The average cost of a certain type of grass seed is $4.00 per box. The standard deviation is $0.10. Using Chebyshev’s theorem, find the minimum percentage of data values that will fall in the range of $3.82 to $4.18.

41. The average U.S. yearly per capita consumption of citrus fruit is 26.8 pounds. Suppose that the distribution of fruit amounts consumed is bell-shaped.
with a standard deviation equal to 4.2 pounds. What percentage of Americans would you expect to consume more than 31 pounds of citrus fruit per year? Source: USDA/Economic Research Service.

42. The average full-time faculty member in a post-secondary degree-granting institution works an average of 53 hours per week.

43. For this data set, find the mean and standard deviation of the variable. The data represent the serum cholesterol levels of 30 individuals. Count the number of data values that fall within 2 standard deviations of the mean. Compare this with the number obtained from Chebyshev’s theorem. Comment on the answer.

211 240 255 219 204
200 212 193 187 205
256 203 210 221 249
231 212 236 204 187
201 247 206 187 200
237 227 221 192 196

44. For this data set, find the mean and standard deviation of the variable. The data represent the ages of 30 customers who ordered a product advertised on television. Count the number of data values that fall within 2 standard deviations of the mean. Compare this with the number obtained from Chebyshev’s theorem. Comment on the answer.

42 44 62 35 20
30 56 20 23 41
55 22 31 27 66
21 18 24 42 25
32 50 31 26 36
39 40 18 36 22

45. Using Chebyshev’s theorem, complete the table to find the minimum percentage of data values that fall within k standard deviations of the mean.

<table>
<thead>
<tr>
<th>k</th>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

46. Use this data set: 10, 20, 30, 40, 50.
   a. Find the standard deviation.
   b. Add 5 to each value, and then find the standard deviation.
   c. Subtract 5 from each value and find the standard deviation.
   d. Multiply each value by 5 and find the standard deviation.
   e. Divide each value by 5 and find the standard deviation.
   f. Generalize the results of parts b through e.
   g. Compare these results with those in Exercise 38.

47. The mean deviation is found by using this formula:

\[ \text{Mean deviation} = \frac{\sum |X - \bar{X}|}{n} \]

where

\[ X = \text{value} \]
\[ \bar{X} = \text{mean} \]
\[ n = \text{number of values} \]
\[ | | = \text{absolute value} \]

Find the mean deviation for these data.

5, 9, 10, 11, 12, 15, 18, 20, 22

48. A measure to determine the skewness of a distribution is called the Pearson coefficient of skewness. The formula is

\[ \text{Skewness} = \frac{3(\bar{X} - \text{MD})}{s} \]

The values of the coefficient usually range from -3 to +3. When the distribution is symmetric, the coefficient is zero; when the distribution is positively skewed, it is positive; and when the distribution is negatively skewed, it is negative.

Using the formula, find the coefficient of skewness for each distribution, and describe the shape of the distribution.

a. Mean = 10, median = 8, standard deviation = 3.
b. Mean = 42, median = 45, standard deviation = 4.
c. Mean = 18.6, median = 18.6, standard deviation = 1.5.
d. Mean = 98, median = 97.6, standard deviation = 4.

49. All values of a data set must be within \( s/\sqrt{n - 1} \) of the mean. If a person collected 25 data values that had a mean of 50 and a standard deviation of 3 and you saw that one data value was 67, what would you conclude?
### Measures of Position

In addition to measures of central tendency and measures of variation, there are measures of position or location. These measures include standard scores, percentiles, deciles, and quartiles. They are used to locate the relative position of a data value in the data set. For example, if a value is located at the 80th percentile, it means that 80% of the values fall below it in the distribution and 20% of the values fall above it. The **median** is the value that corresponds to the 50th percentile, since one-half of the values fall below it and one-half of the values fall above it. This section discusses these measures of position.

#### Standard Scores

There is an old saying, “You can’t compare apples and oranges.” But with the use of statistics, it can be done to some extent. Suppose that a student scored 90 on a music test and 45 on an English exam. Direct comparison of raw scores is impossible, since the exams might not be equivalent in terms of number of questions, value of each question, and so on. However, a comparison of a relative standard similar to both can be made. This comparison uses the mean and standard deviation and is called a standard score or *z* score. (We also use *z* scores in later chapters.)

A **z score** or **standard score** for a value is obtained by subtracting the mean from the value and dividing the result by the standard deviation. The symbol for a standard score is *z*. The formula is

\[
z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}
\]