Use of Standard Deviation

By using the mean and standard deviation, we can find the proportion or percentage of the total observations that fall within a given interval about the mean. This section briefly discusses Chebyshev's theorem and the empirical rule, both of which demonstrate this use of the standard deviation.

Chebyshev's Theorem:

Chebyshev's theorem gives a lower bound for the area under a curve between two points that are on opposite sides of the mean and at the same distance from the mean.

Chebyshev's Theorem For any number k greater than 1, at least $(1 - 1/k^2)$ of the data values lie within k standard deviations of the mean.

Note that k gives the distance between the mean and a point in terms of the number of standard deviations. To find the area under a curve between two points using Chebyshev's theorem, we perform the following steps: First we find the distance between the mean and each of the two given points. Note that both these distances must be the same to apply Chebyshev's theorem. Then we divide the distance calculated above by the standard deviation. This gives the value of k. Finally we substitute the value of k in the formula $1 - \frac{1}{k^2}$ and simplify. Multiply this answer by 100 to obtain the percentage. This gives the minimum area (in percent) under the distribution curve between the two points.

Figure 3.5 illustrates Chebyshev's theorem.

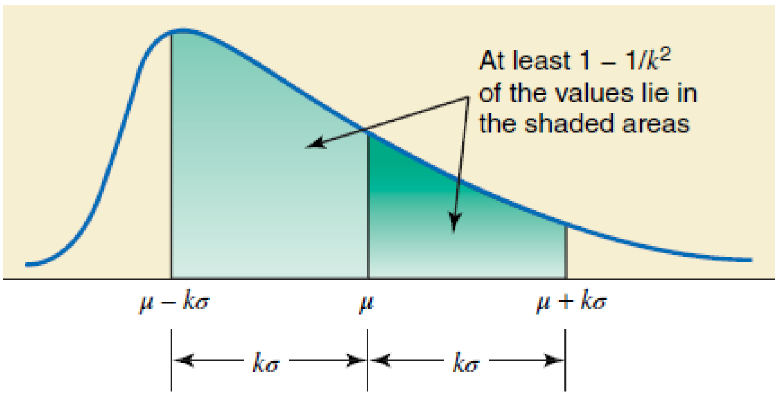


Figure 3.5 Chebyshev's theorem.

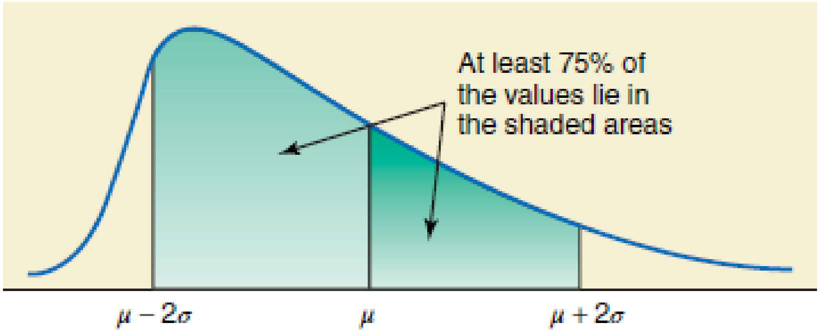
Table 3.14 lists the areas under a distribution curve for different values of k using Chebyshev's theorem.

Table 3.14 Areas Under the Distribution Curve Using Chebyshev's Theorem

k	Interval	$1 - \frac{1}{k^2}$	Minimum Area Within k Standard Deviations
1.5	$\mu \pm 1.5\sigma$	$1 - \frac{1}{1.5^2} = 1 - .44 = .56$	56%
2.0	$\mu \pm 2\sigma$	$1 - \frac{1}{2^2} = 1 - .25 = .75$	75%
2.5	$\mu \pm 2.5\sigma$	$1 - \frac{1}{2.5^2} = 1 - .16 = .84$	84%
3.0	$\mu \pm 3\sigma$	$1 - \frac{1}{3.0^2} = 1 - .11 = .89$	89%

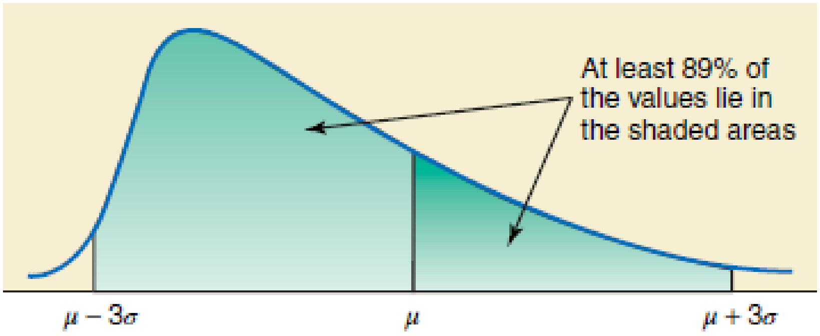
For example, using Table 3.14, if $k = 2$, then according to Chebyshev's theorem, at least .75, or 75%, of the values of a data set lie within two standard deviations of the mean. This is shown in Figure 3.6.

Figure 3.6 Percentage of values within two standard deviations of the mean for Chebyshev's theorem.



If $k = 3$, then according to Chebyshev's theorem, at least .89, or 89%, of the values fall within three standard deviations of the mean. This is shown in Figure 3.7.

Figure 3.7 Percentage of values within three standard deviations of the mean for Chebyshev's theorem.



Although in Figures 3.5 through 3.7 we have used the population notation for the mean and standard deviation, the theorem applies to both sample and population data. Note that Chebyshev's theorem is applicable to a distribution of any shape. However, Chebyshev's theorem can be used only for $k > 1$. This is so because when $k = 1$, the value of $1 - 1/k^2$ is zero, and when $k < 1$, the value of $1 - 1/k^2$ is negative.

<u>Example:</u> Blood Pressure of Women

The average systolic blood pressure for 4000 women who were screened for high blood pressure was found to be 187 mm Hg with a standard deviation of 22. Using Chebyshev's theorem, find the minimum percentage of women in this group who have a systolic blood pressure between 143 and 231 mm Hg.

Solution Let μ and σ be the mean and the standard deviation, respectively, of the systolic blood pressures of these women. Then, from the given information,

$$\mu = 187 \quad \text{and} \quad \sigma = 22$$

To find the percentage of women whose systolic blood pressures are between 143 and 231 mm Hg, the first step is to determine k. As shown below, each of the two points, 143 and 231, is 44 units away from the mean.

$$\begin{array}{ccc} \vert\longleftarrow 44 \longrightarrow\vert & & \vert\longleftarrow 44 \longrightarrow\vert \\ 143 & \mu = 187 & 231 \end{array}$$

The value of k is obtained by dividing the distance between the mean and each point by the standard deviation. Thus,

$$k = 44/22 = 2$$

From Table 3.14, for $k = 2$, the area under the curve is at least 75%. Hence, according to Chebyshev's theorem, at least 75% of the women have a systolic blood pressure between 143 and 231 mm Hg. This percentage is shown in Figure 3.8.

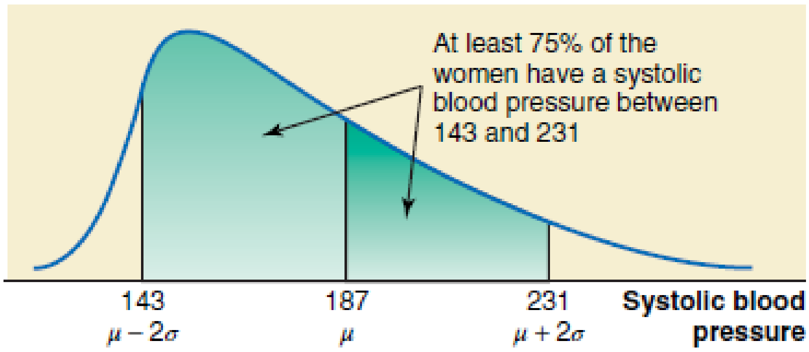


Figure 3.8 Percentage of women with systolic blood pressure between 143 and 231.

Empirical Rule:

Whereas Chebyshev's theorem is applicable to a distribution of any shape, the empirical rule applies only to a specific shape of distribution called a bell-shaped distribution, as shown in Figure 3.9.

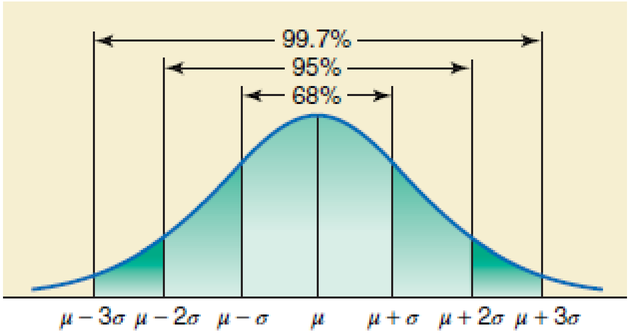


Figure 3.9 Illustration of the empirical rule.

Empirical Rule For a bell-shaped distribution, approximately

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

Table 3.15 lists the areas under a bell-shaped distribution for the above mentioned three intervals.

Table 3.15 Approximate Areas under a Bell-Shaped Distribution Using the Empirical Rule

Interval	Approximate Area
$\mu \pm 1\sigma$	68%
$\mu \pm 2\sigma$	95%
$\mu \pm 3\sigma$	99.7%

Example: Age Distribution of Persons

The age distribution of a sample of 5000 persons is bell-shaped with a mean of 40 years and a standard deviation of 12 years. Determine the approximate percentage of people who are 16 to 64 years old.

Solution We use the empirical rule to find the required percentage because the distribution of ages follows a bell-shaped curve. From the given information, for this distribution,

$$\bar{x} = 40 \text{ years} \quad \text{and} \quad s = 12 \text{ years}$$

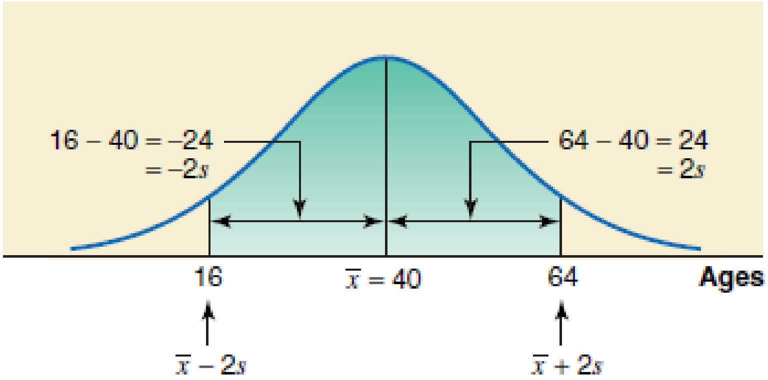


Figure 3.10 Percentage of people who are 16 to 64 years old.

Each of the two points, 16 and 64, is 24 units away from the mean. Dividing 24 by 12, we convert the distance between each of the two points and the mean in terms of the number of standard deviations. Thus, the distance between 16 and 40 and that between 40 and 64 is each equal to $2s$. Consequently, as shown in Figure 3.10, the area from 16 to 64 is the area from $\bar{x} - 2s$ to $\bar{x} + 2s$.

Because, from Table 3.15, the area within two standard deviations of the mean is approximately 95% for a bell-shaped curve, approximately **95%** of the people in the sample are 16 to 64 years old. ■

Practice Problems from text book

3.56 A sample of 2000 observations has a mean of 74 and a standard deviation of 12. Using Chebyshev's theorem, find the minimum percentage of the observations that fall in the intervals $\bar{x} \pm 2s$, $\bar{x} \pm 2.5s$, and $\bar{x} \pm 3s$. Note that $\bar{x} \pm 2s$ represents the interval $\bar{x} - 2s$ to $\bar{x} + 2s$, and so on.

3.57 A large population has a mean of 230 and a standard deviation of 41. Using Chebyshev's theorem, find the minimum percentage of the observations that fall in the intervals $\mu \pm 2\sigma$, $\mu \pm 2.5\sigma$, and $\mu \pm 3\sigma$.

3.58 A large population has a bell-shaped distribution with a mean of 310 and a standard deviation of 37. Using the empirical rule, find the approximate percentage of the observations that fall in the intervals $\mu \pm 1\sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$.

3.59 A sample of 3000 observations has a bell-shaped distribution with a mean of 82 and a standard deviation of 16. Using the empirical rule, find the approximate percentage of the observations that fall in the intervals $\bar{x} \pm 1s$, $\bar{x} \pm 2s$, and $\bar{x} \pm 3s$.

3.60 The mean time taken by all participants to run a road race was found to be 220 minutes with a standard deviation of 20 minutes. Using Chebyshev's theorem, find the minimum percentage of runners who completed the race in

a. 180 to 260 minutes
b. 160 to 280 minutes
c. 170 to 270 minutes

3.61 The one-way commuting times from home to work for all employees working at a large company have a mean of 34 minutes and a standard deviation of 8 minutes.

a. Using Chebyshev's theorem, find the minimum percentage of employees at this company who have one-way commuting times in the following intervals.

i. 14 to 54 minutes **ii.** 18 to 50 minutes

***b.** Using Chebyshev's theorem, find the interval that contains one-way commuting times of at least 89% of the employees at this company.

3.62 The mean monthly mortgage paid by all home owners in a town is \$2365 with a standard deviation of \$340.

a. Using Chebyshev's theorem, find the minimum percentage of all home owners in this town who pay a monthly mortgage of

i. \$1685 to \$3045 **ii.** \$1345 to \$3385

***b.** Using Chebyshev's theorem, find the interval that contains the monthly mortgage payments of at least 84% of all home owners in this town.

3.63 The one-way commuting times from home to work for all employees working at a large company have a bell-shaped curve with a mean of 34 minutes and a standard deviation of 8 minutes. Using the empirical rule, find the approximate percentages of the employees at this company who have one-way commuting times in the following intervals.

a. 10 to 58 minutes **b.** 26 to 42 minutes **c.** 18 to 50 minutes

3.64 The prices of all college textbooks follow a bell-shaped distribution with a mean of \$180 and a standard deviation of \$30.

a. Using the empirical rule, find the (approximate) percentage of all college textbooks with their prices between

i. \$150 and \$210 **ii.** \$120 and \$240

***b.** Using the empirical rule, find the interval that contains the prices of (approximate) 99.7% of college textbooks.