C H A P T E R 6

The Normal Distribution

Objectives

After completing this chapter, you should be able to

1. Identify distributions as symmetric or skewed.
2. Identify the properties of a normal distribution.
3. Find the area under the standard normal distribution, given various *z* values.
4. Find probabilities for a normally distributed variable by transforming it into a standard normal variable.
5. Find specific data values for given percentages, using the standard normal distribution.
6. Use the central limit theorem to solve problems involving sample means for large samples.
7. Use the normal approximation to compute probabilities for a binomial variable.

Outline

Statistics Today

What Is Normal?

Medical researchers have determined so-called normal intervals for a person's blood pressure, cholesterol, triglycerides, and the like. For example, the normal range of systolic blood pressure is 110 to 140. The normal interval for a person's triglycerides is from 30 to 200 milligrams per deciliter (mg/dl). By measuring these variables, a physician can determine if a patient's vital statistics are within the normal interval or if some type of treatment is needed to correct a condition and avoid future illnesses. The question then is, How does one determine the so-called normal intervals? See Statistics Today—Revisited at the end of the chapter.

In this chapter, you will learn how researchers determine normal intervals for specific medical tests by using a normal distribution. You will see how the same methods are used to determine the lifetimes of batteries, the strength of ropes, and many other traits.

Introduction

Historical Note

The name *normal curve* was used by several statisticians, namely, Francis Galton, Charles Sanders, Wilhelm Lexis, and Karl Pearson near the end of the 19th century.

Random variables can be either discrete or continuous. Discrete variables and their distributions were explained in Chapter 5. Recall that a discrete variable cannot assume all values between any two given values of the variables. On the other hand, a continuous variable can assume all values between any two given values of the variables. Examples of continuous variables are the heights of adult men, body temperatures of rats, and cholesterol levels of adults. Many continuous variables, such as the examples just mentioned, have distributions that are bell-shaped, and these are called *approximately normally distributed variables.* For example, if a researcher selects a random sample of 100 adult women, measures their heights, and constructs a histogram, the researcher gets a graph similar to the one shown in Figure 6–1(a). Now, if the researcher increases the sample size and decreases the width of the classes, the histograms will look like the ones shown in Figure 6–1(b) and (c). Finally, if it were possible to measure exactly the heights of all adult females in the United States and plot them, the histogram would approach what is called a *normal distribution,* shown in Figure 6–1(d). This distribution is also known as

Figure 6–1

Histograms for the Distribution of Heights of Adult Women

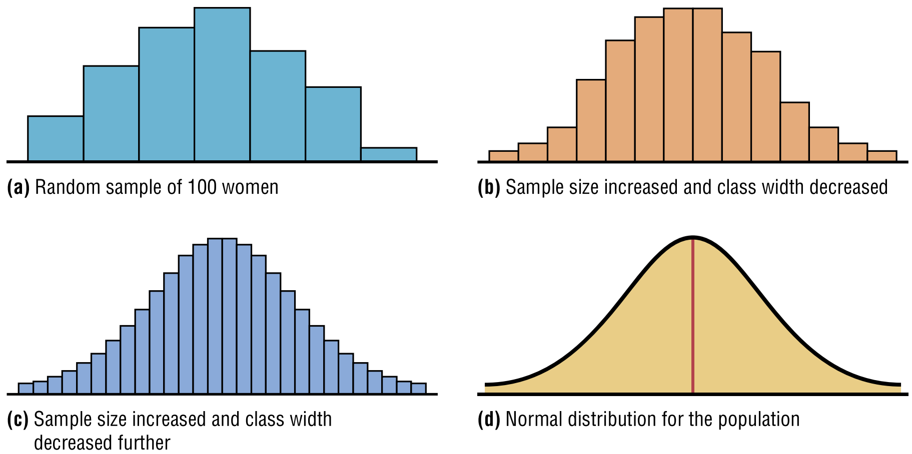
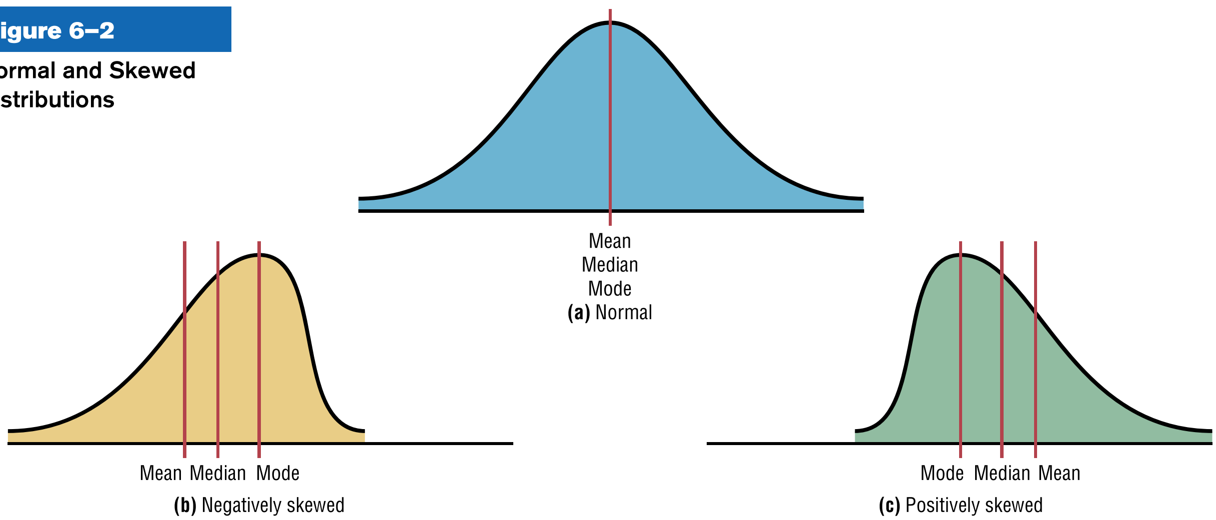


Figure 6–2

Normal and Skewed Distributions



a *bell curve* or a *Gaussian distribution,* named for the German mathematician Carl Friedrich Gauss (1777–1855), who derived its equation.

No variable fits a normal distribution perfectly, since a normal distribution is a theoretical distribution. However, a normal distribution can be used to describe many variables, because the deviations from a normal distribution are very small. This concept will be explained further in Section 6–1.

Objective 1

Identify distributions as symmetric or skewed.

When the data values are evenly distributed about the mean, a distribution is said to be a **symmetric distribution.** (A normal distribution is symmetric.) Figure 6–2(a) shows a symmetric distribution. When the majority of the data values fall to the left or right of the mean, the distribution is said to be *skewed.* When the majority of the data values fall to the right of the mean, the distribution is said to be a **negatively or left-skewed distribution.** The mean is to the left of the median, and the mean and the median are to the left of the mode. See Figure 6–2(b). When the majority of the data values fall to the left of the mean, a distribution is said to be a **positively or right-skewed distribution.** The mean falls to the right of the median, and both the mean and the median fall to the right of the mode. See Figure 6–2(c).

The "tail" of the curve indicates the direction of skewness (right is positive, left is negative). These distributions can be compared with the ones shown in Figure 3–1 in Chapter 3. Both types follow the same principles.

This chapter will present the properties of a normal distribution and discuss its applications. Then a very important fact about a normal distribution called the *central limit theorem* will be explained. Finally, the chapter will explain how a normal distribution curve can be used as an approximation to other distributions, such as the binomial distribution. Since a binomial distribution is a discrete distribution, a correction for continuity may be employed when a normal distribution is used for its approximation.

6–1

Normal Distributions

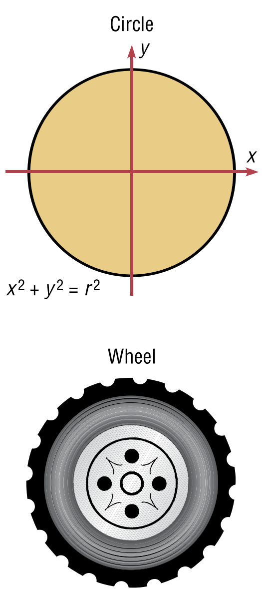
Objective 2

Identify the properties of a normal distribution.

In mathematics, curves can be represented by equations. For example, the equation of the circle shown in Figure 6–3 is $x^2 + y^2 = r^2$, where r is the radius. A circle can be used to represent many physical objects, such as a wheel or a gear. Even though it is not possible to manufacture a wheel that is perfectly round, the equation and the properties of a circle can be used to study many aspects of the wheel, such as area, velocity, and acceleration. In a similar manner, the theoretical curve, called a *normal distribution curve,* can be used to study many variables that are not perfectly normally distributed but are nevertheless approximately normal.

Figure 6–3

Graph of a Circle and an Application



The mathematical equation for a normal distribution is

$$y = \frac{e^{-(X-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

where

$e \approx 2.718$ ($\approx$ means "is approximately equal to")

$\pi \approx 3.14$

μ = population mean

σ = population standard deviation

This equation may look formidable, but in applied statistics, tables or technology is used for specific problems instead of the equation.

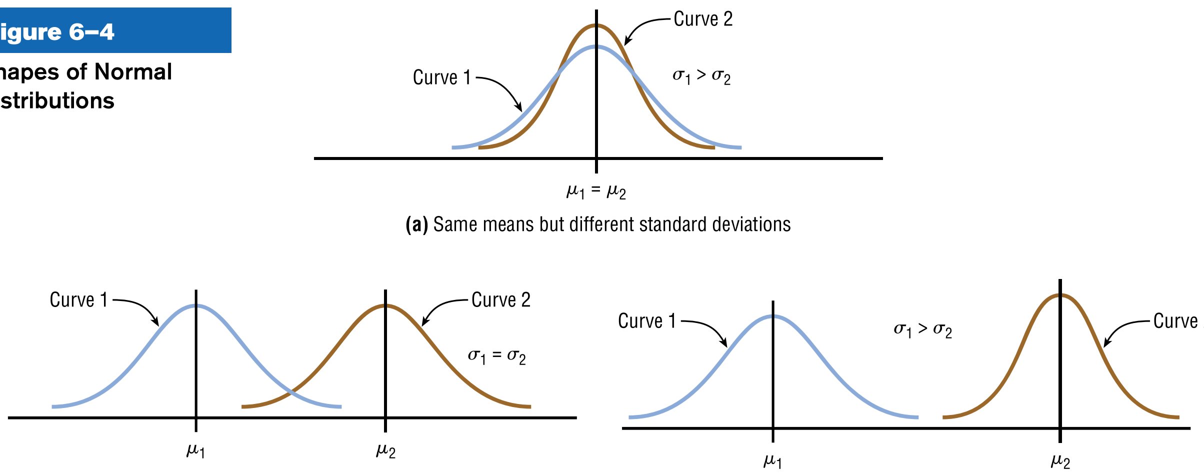
Another important consideration in applied statistics is that the area under a normal distribution curve is used more often than the values on the y axis. Therefore, when a normal distribution is pictured, the y axis is sometimes omitted.

Circles can be different sizes, depending on their diameters (or radii), and can be used to represent wheels of different sizes. Likewise, normal curves have different shapes and can be used to represent different variables.

The shape and position of a normal distribution curve depend on two parameters, the *mean* and the *standard deviation.* Each normally distributed variable has its own normal distribution curve, which depends on the values of the variable's mean and standard deviation. Figure 6–4(a) shows two normal distributions with the same mean values but different standard deviations. The larger the standard deviation, the more dispersed, or spread out, the distribution is. Figure 6–4(b) shows two normal distributions with the same standard deviation but with different means. These curves have the same shapes but are located at different positions on the x axis. Figure 6–4(c) shows two normal distributions with different means and different standard deviations.

Figure 6–4

Shapes of Normal Distributions



(a) Same means but different standard deviations

(b) Different means but same standard deviations

(c) Different means and different standard deviations

Historical Notes

The discovery of the equation for a normal distribution can be traced to three mathematicians. In 1733, the French mathematician Abraham DeMoivre derived an equation for a normal distribution based on the random variation of the number of heads appearing when a large number of coins were tossed. Not realizing any connection with the naturally occurring variables, he showed this formula to only a few friends. About 100 years later, two mathematicians, Pierre Laplace in France and Carl Gauss in Germany, derived the equation of the normal curve independently and without any knowledge of DeMoivre's work. In 1924, Karl Pearson found that DeMoivre had discovered the formula before Laplace or Gauss.

A **normal distribution** is a continuous, symmetric, bell-shaped distribution of a variable.

The properties of a normal distribution, including those mentioned in the definition, are explained next.

Summary of the Properties of the Theoretical Normal Distribution

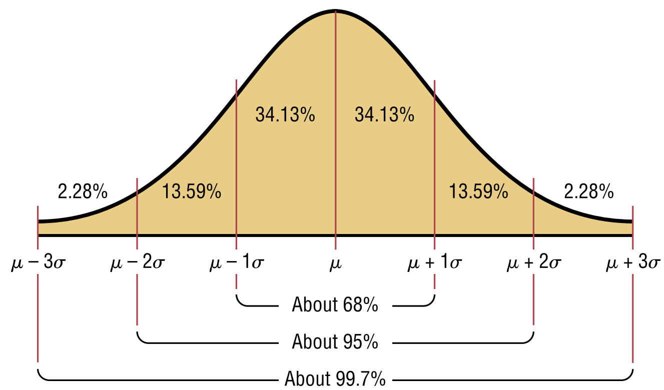
1. A normal distribution curve is bell-shaped.
2. The mean, median, and mode are equal and are located at the center of the distribution.
3. A normal distribution curve is unimodal (i.e., it has only one mode).
4. The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.
5. The curve is continuous; that is, there are no gaps or holes. For each value of X, there is a corresponding value of Y.
6. The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis—but it gets increasingly closer.
7. The total area under a normal distribution curve is equal to 1.00, or 100%. This fact may seem unusual, since the curve never touches the x axis, but one can prove it mathematically by using calculus. (The proof is beyond the scope of this textbook.)
8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%. See Figure 6–5, which also shows the area in each region.

The values given in item 8 of the summary follow the *empirical rule* for data given in Section 3–2.

You must know these properties in order to solve problems involving distributions that are approximately normal.

Figure 6–5

Areas Under a Normal Distribution Curve



The Standard Normal Distribution

Since each normally distributed variable has its own mean and standard deviation, as stated earlier, the shape and location of these curves will vary. In practical applications, then, you would have to have a table of areas under the curve for each variable. To simplify this situation, statisticians use what is called the *standard normal distribution.*

Objective 3

Find the area under the standard normal distribution, given various z values.

> The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.

The standard normal distribution is shown in Figure 6–6.

The values under the curve indicate the proportion of area in each section. For example, the area between the mean and 1 standard deviation above or below the mean is about 0.3413, or 34.13%.

The formula for the standard normal distribution is

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

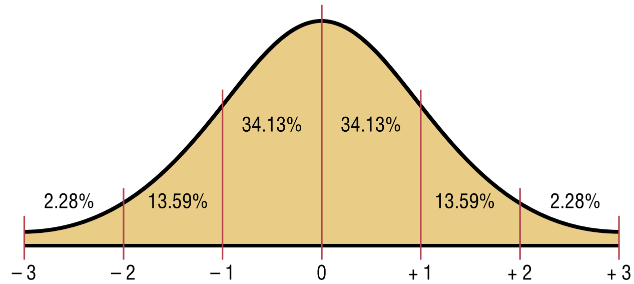
$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

This is the same formula used in Section 3–3. The use of this formula will be explained in Section 6–3.

As stated earlier, the area under a normal distribution curve is used to solve practical application problems, such as finding the percentage of adult women whose height is between 5 feet 4 inches and 5 feet 7 inches, or finding the probability that a new battery will last longer than 4 years. Hence, the major emphasis of this section will be to show the procedure for finding the area under the standard normal distribution curve for any z value. The applications will be shown in Section 6–2. Once the X values are transformed by using the preceding formula, they are called z values. The ***z* value** or ***z* score** is actually the number of standard deviations that a particular X value is away from the mean. Table E in Appendix C gives the area (to four decimal places) under the standard normal curve for any z value from −3.49 to 3.49.

Figure 6–6

Standard Normal Distribution



Interesting Fact

Bell-shaped distributions occurred quite often in early coin-tossing and die-rolling experiments.

Finding Areas Under the Standard Normal Distribution Curve

For the solution of problems using the standard normal distribution, a two-step process is recommended with the use of the Procedure Table shown.

The two steps are

Step 1 Draw the normal distribution curve and shade the area.

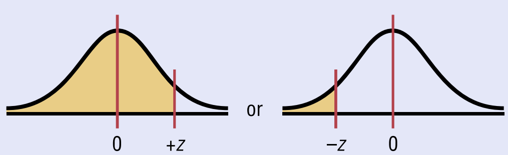
Step 2 Find the appropriate figure in the Procedure Table and follow the directions given.

There are three basic types of problems, and all three are summarized in the Procedure Table. Note that this table is presented as an aid in understanding how to use the standard normal distribution table and in visualizing the problems. After learning the procedures, you should not find it necessary to refer to the Procedure Table for every problem.

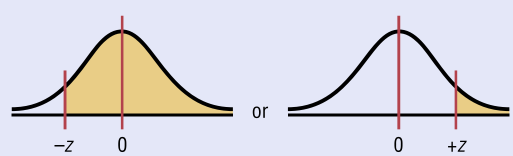
Procedure Table

Finding the Area Under the Standard Normal Distribution Curve

1. To the left of any z value:
   Look up the z value in the table and use the area given.



2. To the right of any z value:
   Look up the z value and subtract the area from 1.



3. Between any two z values:
   Look up both z values and subtract the corresponding areas.

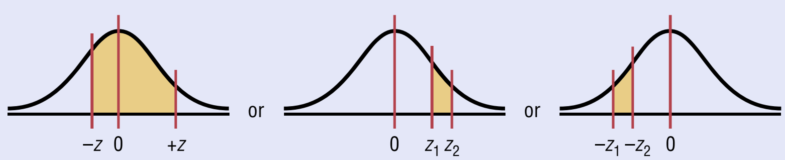


Figure 6–7

Table E Area Value for $z = 1.39$

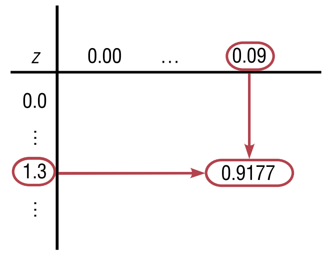


Table E in Appendix C gives the area under the normal distribution curve to the left of any z value given in two decimal places. For example, the area to the left of a z value of 1.39 is found by looking up 1.3 in the left column and 0.09 in the top row. Where the two lines meet gives an area of 0.9177. See Figure 6–7.

Example 6–1

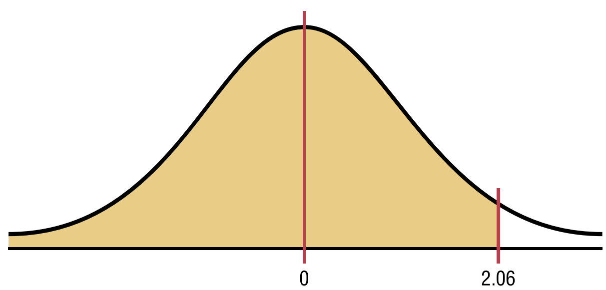
Find the area to the left of $z = 2.06$.

Solution

Step 1 Draw the figure. The desired area is shown in Figure 6–8.

Figure 6–8

Area Under the Standard Normal Distribution Curve for Example 6–1



Step 2 We are looking for the area under the standard normal distribution to the left of $z = 2.06$. Since this is an example of the first case, look up the area in the table. It is 0.9803. Hence, 98.03% of the area is less than $z = 2.06$.

Example 6–2

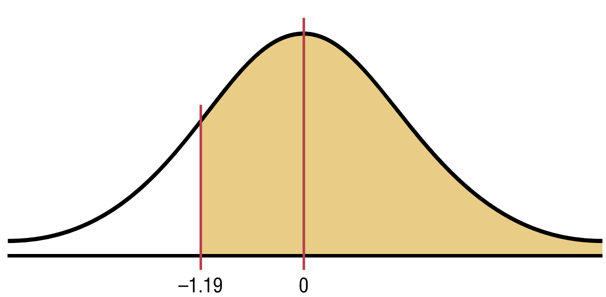
Find the area to the right of $z = -1.19$.

Solution

Step 1 Draw the figure. The desired area is shown in Figure 6–9.

Figure 6–9

Area Under the Standard Normal Distribution Curve for Example 6–2



Step 2 We are looking for the area to the right of $z = -1.19$. This is an example of the second case. Look up the area for $z = -1.19$. It is 0.1170. Subtract it from 1.0000. $1.0000 - 0.1170 = 0.8830$. Hence, 88.30% of the area under the standard normal distribution curve is to the left of $z = -1.19$.

Example 6–3

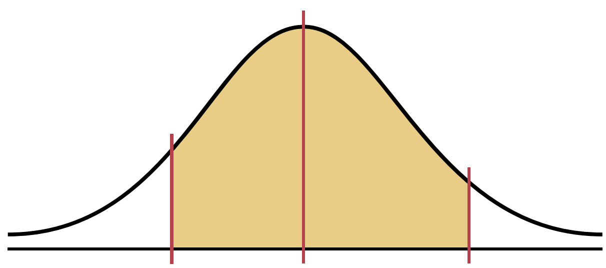
Find the area between $z = +1.68$ and $z = -1.37$.

Solution

Step 1 Draw the figure as shown. The desired area is shown in Figure 6–10.

Figure 6–10

Area Under the Standard Normal Distribution Curve for Example 6–3



Step 2 Since the area desired is between two given z values, look up the areas corresponding to the two z values and subtract the smaller area from the larger area. (Do not subtract the z values.) The area for $z = +1.68$ is 0.9535, and the area for $z = -1.37$ is 0.0853. The area between the two z values is $0.9535 - 0.0853 = 0.8682$ or 86.82%.

A Normal Distribution Curve as a Probability Distribution Curve

A normal distribution curve can be used as a probability distribution curve for normally distributed variables. Recall that a normal distribution is a *continuous distribution,* as opposed to a discrete probability distribution, as explained in Chapter 5. The fact that it is continuous means that there are no gaps in the curve. In other words, for every z value on the x axis, there is a corresponding height, or frequency, value.

The area under the standard normal distribution curve can also be thought of as a probability. That is, if it were possible to select any z value at random, the probability of choosing one, say, between 0 and 2.00 would be the same as the area under the curve between 0 and 2.00. In this case, the area is 0.4772. Therefore, the probability of randomly selecting any z value between 0 and 2.00 is 0.4772. The problems involving probability are solved in the same manner as the previous examples involving areas in this section. For example, if the problem is to find the probability of selecting a z value between 2.25 and 2.94, solve it by using the method shown in case 3 of the Procedure Table.

For probabilities, a special notation is used. For example, if the problem is to find the probability of any z value between 0 and 2.32, this probability is written as $P(0 < z < 2.32)$.

Note: In a continuous distribution, the probability of any exact z value is 0 since the area would be represented by a vertical line above the value. But vertical lines in theory have no area. So $P(a \leq z \leq b) = P(a < z < b)$.

Example 6–4

Find the probability for each.

a. $P(0 < z < 2.32)$

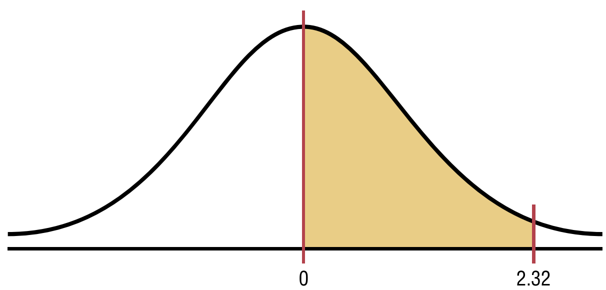
b. $P(z < 1.65)$

c. $P(z > 1.91)$

Solution

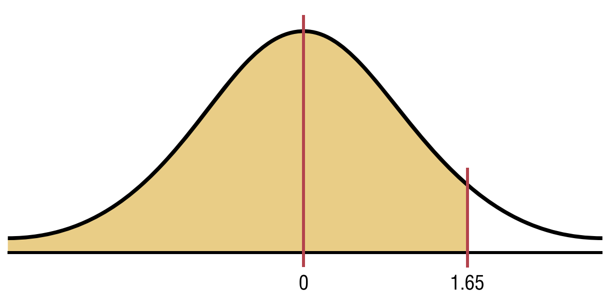
a. $P(0 < z < 2.32)$ means to find the area under the standard normal distribution curve between 0 and 2.32. First look up the area corresponding to 2.32. It is 0.9898. Then look up the area corresponding to $z = 0$. It is 0.500. Subtract the two areas: $0.9898 - 0.5000 = 0.4898$. Hence the probability is 0.4898, or 48.98%. This is shown in Figure 6–11.

Figure 6–11
Area Under the Standard Normal Distribution Curve for Part *a* of Example 6–4



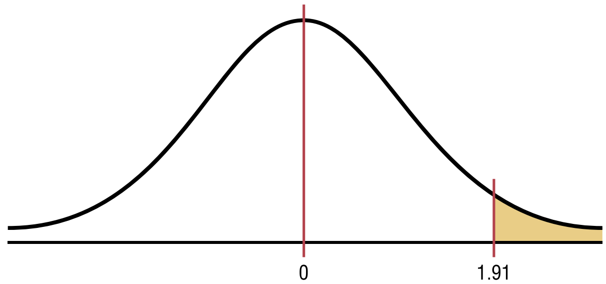
b. $P(z < 1.65)$ is represented in Figure 6–12. Look up the area corresponding to $z = 1.65$ in Table E. It is 0.9505. Hence, $P(z < 1.65) = 0.9505$, or 95.05%.

Figure 6–12
Area Under the Standard Normal Distribution Curve for Part *b* of Example 6–4



c. $P(z > 1.91)$ is shown in Figure 6–13. Look up the area that corresponds to $z = 1.91$. It is 0.9719. Then subtract this area from 1.0000. $P(z > 1.91) = 1.0000 - 0.9719 = 0.0281$, or 2.81%.

Figure 6–13
Area Under the Standard Normal Distribution Curve for Part *c* of Example 6–4



Sometimes, one must find a specific *z* value for a given area under the standard normal distribution curve. The procedure is to work backward, using Table E.

Since Table E is cumulative, it is necessary to locate the cumulative area up to a given *z* value. Example 6–5 shows this.

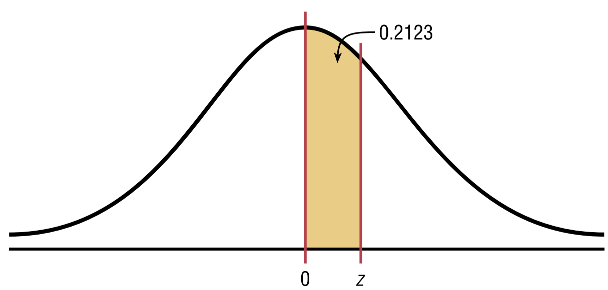
Example 6–5

Find the *z* value such that the area under the standard normal distribution curve between 0 and the *z* value is 0.2123.

Solution

Draw the figure. The area is shown in Figure 6–14.

Figure 6–14
Area Under the Standard Normal Distribution Curve for Example 6–5



In this case it is necessary to add 0.5000 to the given area of 0.2123 to get the cumulative area of 0.7123. Look up the area in Table E. The value in the left column is 0.5, and the top value is 0.06. Add these two values to get $z = 0.56$. See Figure 6–15.

Figure 6–15
Finding the *z* Value from Table E for Example 6–5

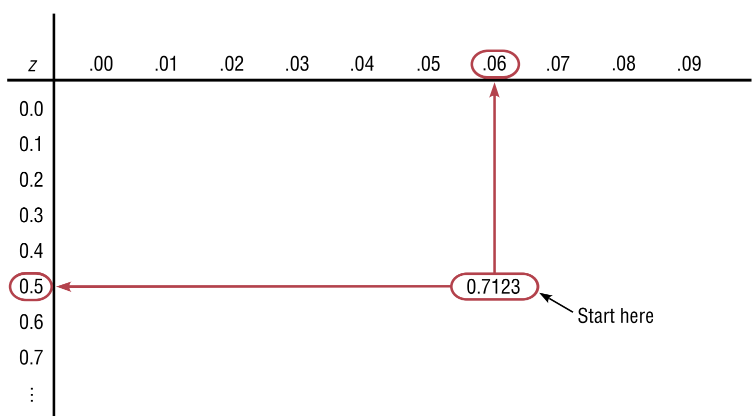
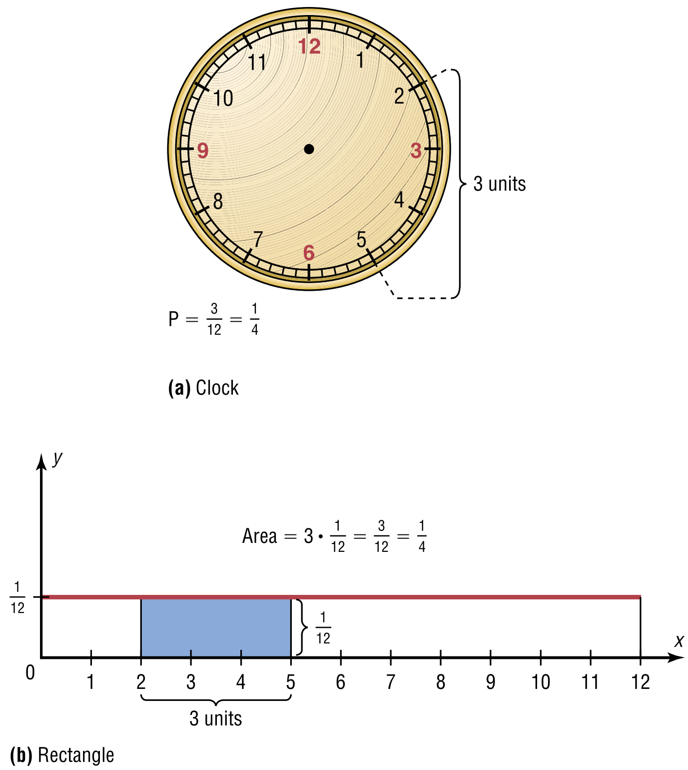


Figure 6–16

The Relationship Between Area and Probability



If the exact area cannot be found, use the closest value. For example, if you wanted to find the *z* value for an area 0.9241, the closest area is 0.9236, which gives a *z* value of 1.43. See Table E in Appendix C.

The rationale for using an area under a continuous curve to determine a probability can be understood by considering the example of a watch that is powered by a battery. When the battery goes dead, what is the probability that the minute hand will stop somewhere between the numbers 2 and 5 on the face of the watch? In this case, the values of the variable constitute a continuous variable since the hour hand can stop anywhere on the dial's face between 0 and 12 (one revolution of the minute hand). Hence, the sample space can be considered to be 12 units long, and the distance between the numbers 2 and 5 is $5 - 2$, or 3 units. Hence, the probability that the minute hand stops on a number between 2 and 5 is $\frac{3}{12} = \frac{1}{4}$. See Figure 6–16(a).

The problem could also be solved by using a graph of a continuous variable. Let us assume that since the watch can stop anytime at random, the values where the minute hand would land are spread evenly over the range of 0 through 12. The graph would then consist of a *continuous uniform distribution* with a range of 12 units. Now if we require the area under the curve to be 1 (like the area under the standard normal distribution), the height of the rectangle formed by the curve and the x axis would need to be $\frac{1}{12}$. The reason is that the area of a rectangle is equal to the base times the height. If the base is 12 units long, then the height has to be $\frac{1}{12}$ since $12 \cdot \frac{1}{12} = 1$.

The area of the rectangle with a base from 2 through 5 would be $3 \cdot \frac{1}{12}$, or $\frac{1}{4}$. See Figure 6–16(b). Notice that the area of the small rectangle is the same as the probability found previously. Hence the area of this rectangle corresponds to the probability of this event. The same reasoning can be applied to the standard normal distribution curve shown in Example 6–5.

Finding the area under the standard normal distribution curve is the first step in solving a wide variety of practical applications in which the variables are normally distributed. Some of these applications will be presented in Section 6–2.

Example XL6–3

Find the area between $z = -2.04$ and $z = 1.99$.
In a blank cell type: =NORMSDIST(1.99) − NORMSDIST(−2.04)
Answer: 0.956029

Finding a *z* value given an area under the standard normal distribution curve

Example XL6–4

Find a *z* score given the cumulative area (area to the left of *z*) is 0.0250.
In a blank cell type: =NORMSINV(.025)
Answer: −1.95996

6–2 Applications of the Normal Distribution

Objective 4

Find probabilities for a normally distributed variable by transforming it into a standard normal variable.

The standard normal distribution curve can be used to solve a wide variety of practical problems. The only requirement is that the variable be normally or approximately normally distributed. There are several mathematical tests to determine whether a variable is normally distributed. See the Critical Thinking Challenges on page 352. For all the problems presented in this chapter, you can assume that the variable is normally or approximately normally distributed.

To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the formula

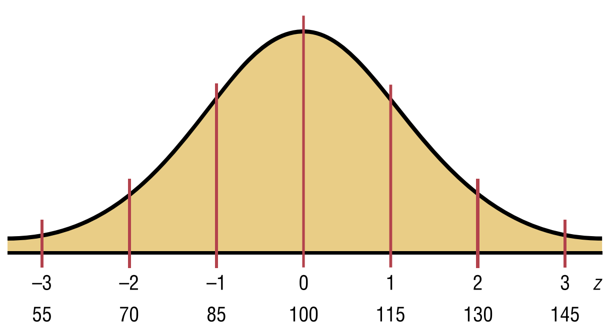
$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \qquad \text{or} \qquad z = \frac{X - \mu}{\sigma}$$

This is the same formula presented in Section 3–3. This formula transforms the values of the variable into standard units or *z* values. Once the variable is transformed, then the Procedure Table and Table E in Appendix C can be used to solve problems.

For example, suppose that the scores for a standardized test are normally distributed, have a mean of 100, and have a standard deviation of 15. When the scores are transformed to *z* values, the two distributions coincide, as shown in Figure 6–17. (Recall that the *z* distribution has a mean of 0 and a standard deviation of 1.)

Figure 6–17

Test Scores and Their Corresponding *z* Values



To solve the application problems in this section, transform the values of the variable to *z* values and then find the areas under the standard normal distribution, as shown in Section 6–1.

Example 6–6

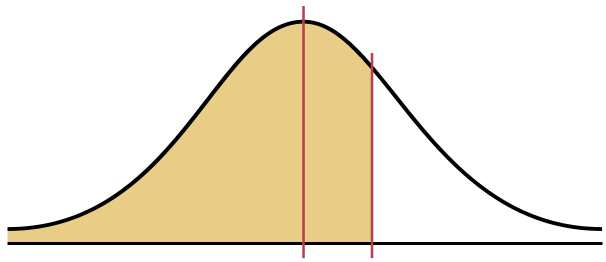
Summer Spending

A survey found that women spend on average \$146.21 on beauty products during the summer months. Assume the standard deviation is \$29.44. Find the percentage of women who spend less than \$160.00. Assume the variable is normally distributed.

Solution

Step 1 Draw the figure and represent the area as shown in Figure 6–18.

Figure 6–18
Area Under a Normal Curve for Example 6–6

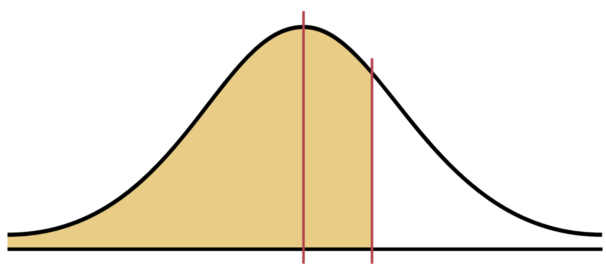


Step 2 Find the z value corresponding to \$160.00.

$$z = \frac{X - \mu}{\sigma} = \frac{\$160.00 - \$146.21}{\$29.44} = 0.47$$

Hence \$160.00 is 0.47 of a standard deviation above the mean of \$146.21, as shown in the z distribution in Figure 6–19.

Figure 6–19
Area and z Values for Example 6–6



Step 3 Find the area, using Table E. The area under the curve to the left of $z = 0.47$ is 0.6808.

Therefore 0.6808, or 68.08%, of the women spend less than \$160.00 on beauty products during the summer months.

Example 6–7

Monthly Newspaper Recycling

Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating

a. Between 27 and 31 pounds per month

b. More than 30.2 pounds per month

Assume the variable is approximately normally distributed.

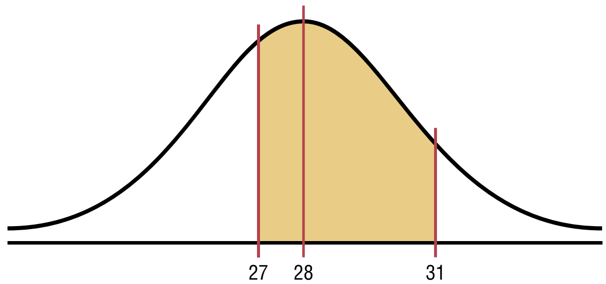
Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

Solution *a*

Step 1 Draw the figure and represent the area. See Figure 6–20.

Figure 6–20

Area Under a Normal Curve for Part *a* of Example 6–7



Historical Note

Astronomers in the late 1700s and the 1800s used the principles underlying the normal distribution to correct measurement errors that occurred in charting the positions of the planets.

Step 2 Find the two z values.

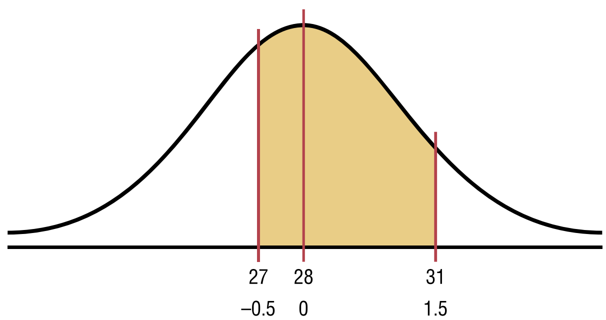
$$z_1 = \frac{X - \mu}{\sigma} = \frac{27 - 28}{2} = -\frac{1}{2} = -0.5$$

$$z_2 = \frac{X - \mu}{\sigma} = \frac{31 - 28}{2} = \frac{3}{2} = 1.5$$

Step 3 Find the appropriate area, using Table E. The area to the left of z_2 is 0.9332, and the area to the left of z_1 is 0.3085. Hence the area between z_1 and z_2 is $0.9332 - 0.3085 = 0.6247$. See Figure 6–21.

Figure 6–21

Area and *z* Values for Part *a* of Example 6–7



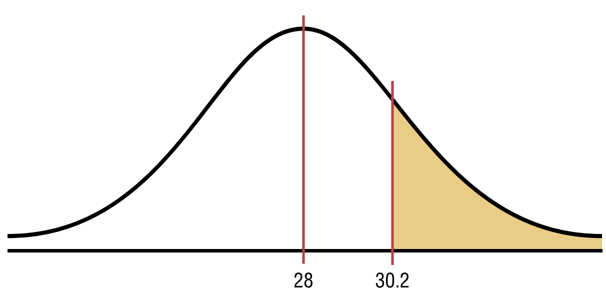
Hence, the probability that a randomly selected household generates between 27 and 31 pounds of newspapers per month is 62.47%.

Solution *b*

Step 1 Draw the figure and represent the area, as shown in Figure 6–22.

Figure 6–22

Area Under a Normal Curve for Part *b* of Example 6–7



Step 2 Find the z value for 30.2.

$$z = \frac{X - \mu}{\sigma} = \frac{30.2 - 28}{2} = \frac{2.2}{2} = 1.1$$

Step 3 Find the appropriate area. The area to the left of $z = 1.1$ is 0.8643. Hence the area to the right of $z = 1.1$ is $1.0000 - 0.8643 = 0.1357$.

Hence, the probability that a randomly selected household will accumulate more than 30.2 pounds of newspapers is 0.1357, or 13.57%.

A normal distribution can also be used to answer questions of "How many?" This application is shown in Example 6–8.

Example 6–8

Coffee Consumption

Americans consume an average of 1.64 cups of coffee per day. Assume the variable is approximately normally distributed with a standard deviation of 0.24 cup. If 500 individuals are selected, approximately how many will drink less than 1 cup of coffee per day?

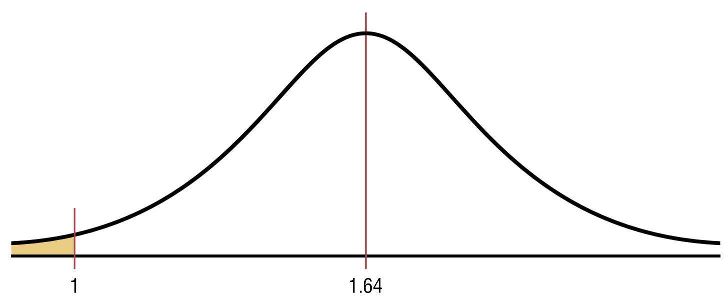
Source: *Chicago Sun-Times.*

Solution

Step 1 Draw a figure and represent the area as shown in Figure 6–23.

Figure 6–23

Area Under a Normal Curve for Example 6–8



Step 2 Find the z value for 1.

$$z = \frac{X - \mu}{\sigma} = \frac{1 - 1.64}{0.24} = -2.67$$

Step 3 Find the area to the left of $z = -2.67$. It is 0.0038.

Step 4 To find how many people drank less than 1 cup of coffee, multiply the sample size 500 by 0.0038 to get 1.9. Since we are asking about people, round the answer to 2 people. Hence, approximately 2 people will drink less than 1 cup of coffee a day.

Note: For problems using percentages, be sure to change the percentage to a decimal before multiplying. Also, round the answer to the nearest whole number, since it is not possible to have 1.9 people.

Finding Data Values Given Specific Probabilities

A normal distribution can also be used to find specific data values for given percentages. This application is shown in Example 6–9.

Example 6–9

Objective 5

Find specific data values for given percentages, using the standard normal distribution.

Police Academy Qualifications

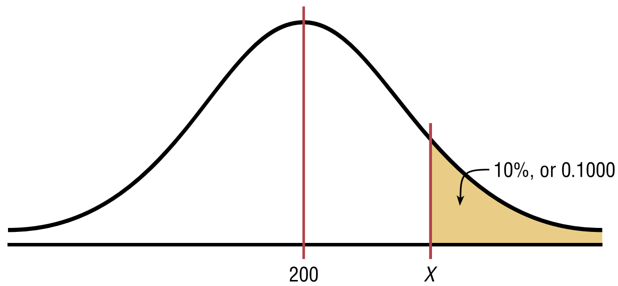
To qualify for a police academy, candidates must score in the top 10% on a general abilities test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume the test scores are normally distributed.

Solution

Since the test scores are normally distributed, the test value X that cuts off the upper 10% of the area under a normal distribution curve is desired. This area is shown in Figure 6–24.

Figure 6–24

Area Under a Normal Curve for Example 6–9



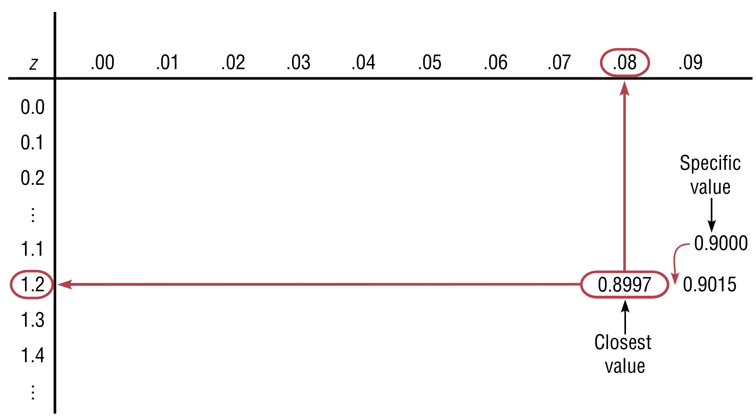
Work backward to solve this problem.

Step 1 Subtract 0.1000 from 1.000 to get the area under the normal distribution to the left of x: $1.0000 - 0.10000 = 0.9000$.

Step 2 Find the z value that corresponds to an area of 0.9000 by looking up 0.9000 in the area portion of Table E. If the specific value cannot be found, use the closest value—in this case 0.8997, as shown in Figure 6–25. The corresponding z value is 1.28. (If the area falls exactly halfway between two z values, use the larger of the two z values. For example, the area 0.9500 falls halfway between 0.9495 and 0.9505. In this case use 1.65 rather than 1.64 for the z value.)

Figure 6–25

Finding the z Value from Table E (Example 6–9)



Step 3 Substitute in the formula $z = (X - \mu)/\sigma$ and solve for X.

$$1.28 = \frac{X - 200}{20}$$

$$(1.28)(20) + 200 = X$$

$$25.60 + 200 = X$$

$$225.60 = X$$

$$226 = X$$

A score of 226 should be used as a cutoff. Anybody scoring 226 or higher qualifies.

Interesting Fact

Americans are the largest consumers of chocolate. We spend $16.6 billion annually.

Instead of using the formula shown in step 3, you can use the formula $X = z \cdot \sigma + \mu$. This is obtained by solving

$$z = \frac{X - \mu}{\sigma}$$

for X as shown.

$z \cdot \sigma = X - \mu$	Multiply both sides by σ.
$z \cdot \sigma + \mu = X$	Add μ to both sides.
$X = z \cdot \sigma + \mu$	Exchange both sides of the equation.

Formula for Finding *X*

When you must find the value of X, you can use the following formula:

$$X = z \cdot \sigma + \mu$$

Example 6–10

Systolic Blood Pressure

For a medical study, a researcher wishes to select people in the middle 60% of the population based on blood pressure. If the mean systolic blood pressure is 120 and the standard deviation is 8, find the upper and lower readings that would qualify people to participate in the study.

Solution

Assume that blood pressure readings are normally distributed; then cutoff points are as shown in Figure 6–26.

Figure 6–26

Area Under a Normal Curve for Example 6–10

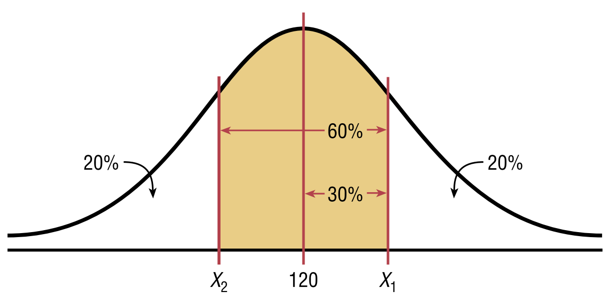


Figure 6–26 shows that two values are needed, one above the mean and one below the mean. To get the area to the left of the positive z value, add 0.5000 + 0.3000 = 0.8000 (30% = 0.3000). The z value with area to the left closest to 0.8000 is 0.84. Substituting in the formula $X = z\sigma + \mu$ gives

$$X_1 = z\sigma + \mu = (0.84)(8) + 120 = 126.72$$

The area to the left of the negative z value is 20%, or 0.2000. The area closest to 0.2000 is −0.84.

$$X_2 = (-0.84)(8) + 120 = 113.28$$

Therefore, the middle 60% will have blood pressure readings of $113.28 < X < 126.72$.

As shown in this section, a normal distribution is a useful tool in answering many questions about variables that are normally or approximately normally distributed.