CHAPTER 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.

6–1 Introduction

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

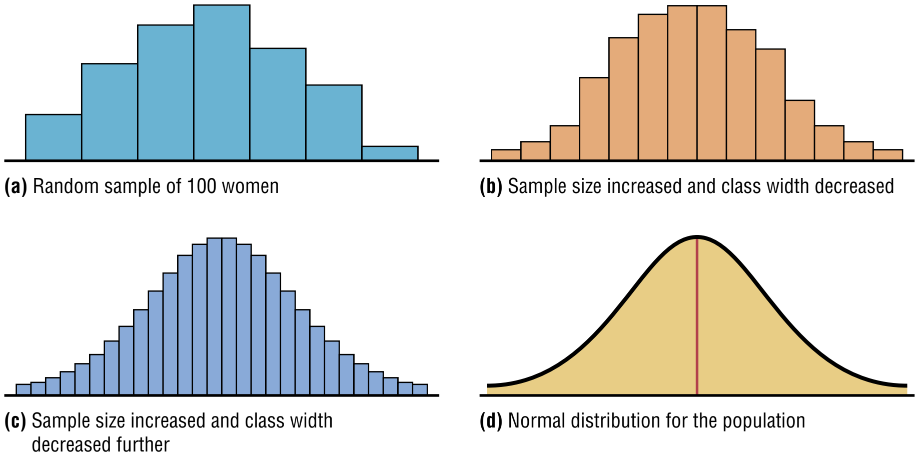
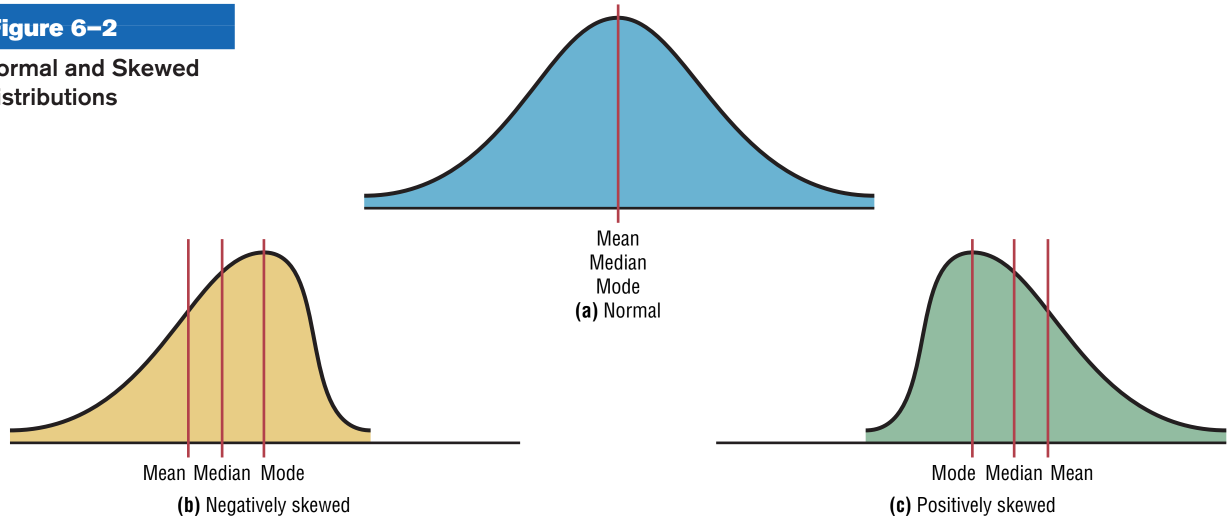


Figure 6–2

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 6–2.

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 negative). These distributions can be compared with the ones shown in Figure 3–1 on page 109. 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.

6–2 Properties of a Normal Distribution

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.

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 6–3

Graph of a Circle and an Application

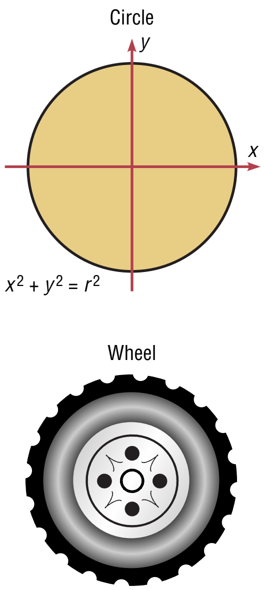
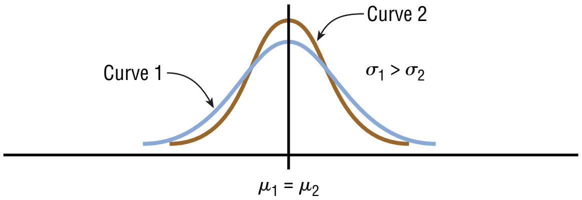
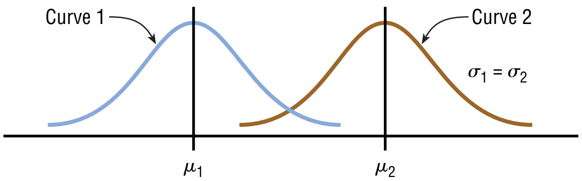


Figure 6–4

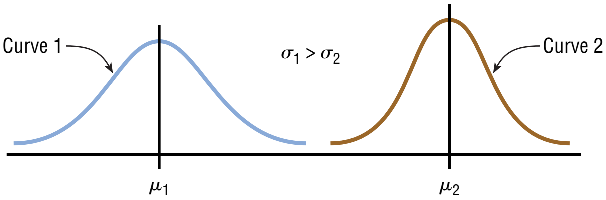
Shapes of Normal Distributions



(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

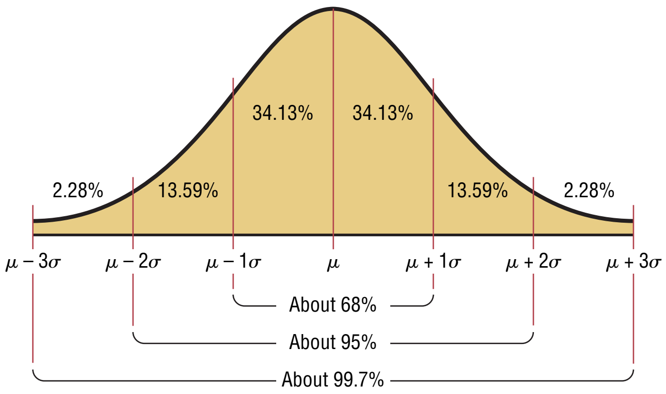
1. A normal distribution curve is bell-shaped.
2. The mean, median, and mode are equal and are located at the center of the distribution.
3. A normal distribution curve is unimodal (i.e., it has only one mode).
4. The curve is symmetric about the mean, which is equivalent to saying that its shape is 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 number 8 of the summary follow the *empirical rule* for data given in Section 3–3.

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

Figure 6–5

Areas Under a Normal Distribution Curve



6–3 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, one 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:

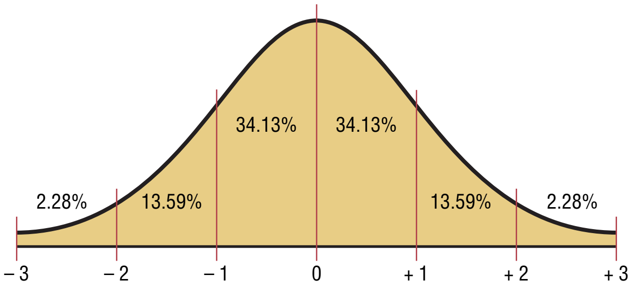
$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \quad \text{or} \quad z = \frac{X - \mu}{\sigma}$$

This is the same formula used in Section 3–4. The use of this formula will be explained in Section 6–4.

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 6–4. Once the X values are transformed by using the preceding formula, they are called z values. The ***z* value** 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 0 to 3.09.

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 four-step procedure is recommended with the use of the Procedure Table shown.

Step 1 Draw a picture.

Step 2 Shade the area desired.

Step 3 Find the correct figure in the following Procedure Table (the figure that is similar to the one you've drawn).

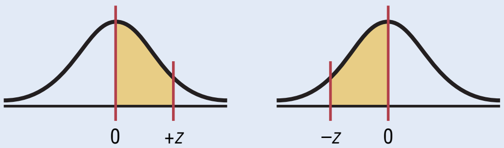
Step 4 Follow the directions given in the appropriate block of the Procedure Table to get the desired area.

There are seven basic types of problems and all seven 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, one should *not* find it necessary to refer to the procedure table for every problem.

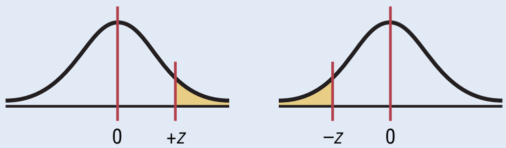
Procedure Table

Finding the Area Under the Standard Normal Distribution Curve

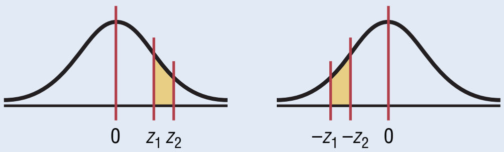
1. Between 0 and any z value:
   Look up the z value in the table to get the area.



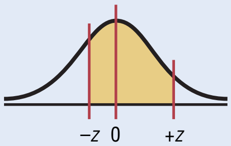
2. In any tail:
   a. Look up the z value to get the area.
   b. Subtract the area from 0.5000.



3. Between two z values on the same side of the mean:
   a. Look up both z values to get the areas.
   b. Subtract the smaller area from the larger area.



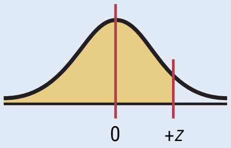
4. Between two z values on opposite sides of the mean:
   a. Look up both z values to get the areas.
   b. Add the areas.



Procedure Table (concluded)

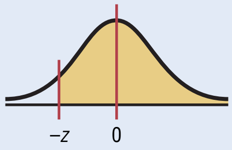
Finding the Area Under the Standard Normal Distribution Curve

5. To the left of any z value, where z is greater than the mean:



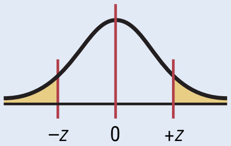
a. Look up the z value to get the area.
b. Add 0.5000 to the area.

6. To the right of any z value, where z is less than the mean:



a. Look up the z value in the table to get the area.
b. Add 0.5000 to the area.

7. In any two tails:



a. Look up the z values in the table to get the areas.
b. Subtract both areas from 0.5000.
c. Add the answers.

Procedure
1. Draw the picture.
2. Shade the area desired.
3. Find the correct figure.
4. Follow the directions.

Note: Table E gives the area between 0 and any z value to the right of 0, and all areas are positive.

Situation 1 Find the area under the standard normal curve between 0 and any z value.

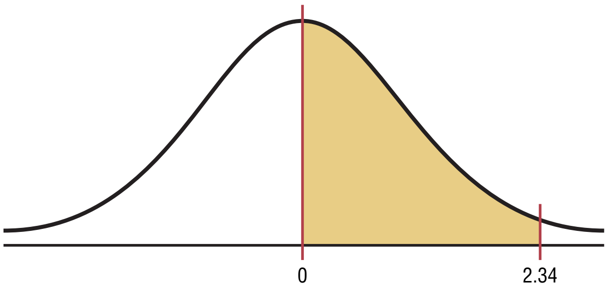
Example 6–1

Find the area under the standard normal distribution curve between $z = 0$ and $z = 2.34$.

Solution

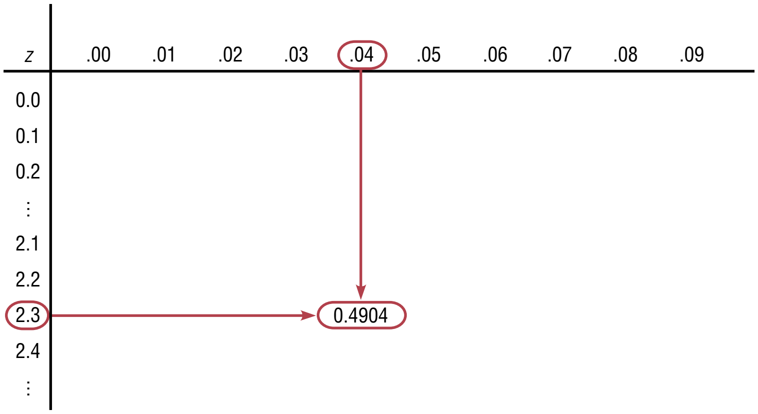
Draw the figure and represent the area as shown in Figure 6–7.

Figure 6–7
Area Under the Standard Normal Distribution Curve for Example 6–1



Since Table E gives the area between 0 and any z value to the right of 0, one need only look up the z value in the table. Find 2.3 in the left column and 0.04 in the top row. The value where the column and row meet in the table is the answer, 0.4904. See Figure 6–8. Hence, the area is 0.4904, or 49.04%.

Figure 6–8
Using Table E in the Appendix for Example 6–1



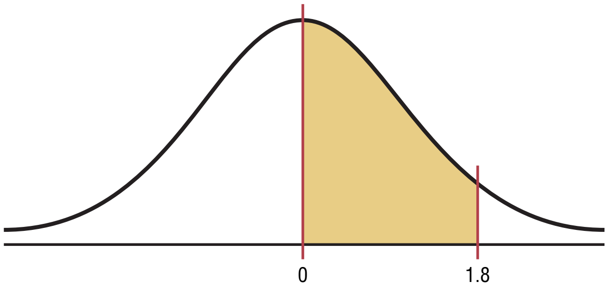
Example 6–2

Find the area between $z = 0$ and $z = 1.8$.

Solution

Draw the figure and represent the area as shown in Figure 6–9.

Figure 6–9
Area Under the Standard Normal Curve for Example 6–2



Find the area in Table E by finding 1.8 in the left column and 0.00 in the top row. The area is 0.4641, or 46.41%.

Next, one must be able to find the areas for values that are not in Table E. This is done by using the properties of the normal distribution described in Section 6–2.

Example 6–3

Find the area between $z = 0$ and $z = -1.75$.

Solution

Represent the area as shown in Figure 6–10.

Figure 6–10
Area Under the Standard Normal Curve for Example 6–3

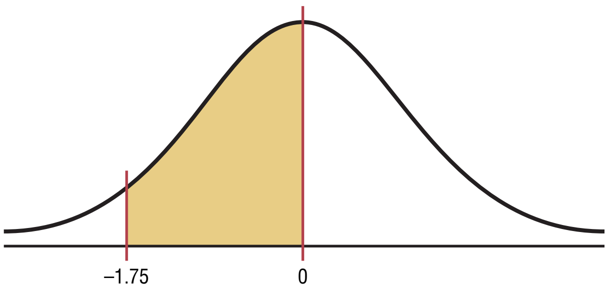


Table E does not give the areas for negative values of z. But since the normal distribution is symmetric about the mean, the area to the left of the mean (in this case, the mean is 0) is the same as the area to the right of the mean. Hence one need only look up the area for $z = +1.75$, which is 0.4599, or 45.99%. This solution is summarized in block 1 in the Procedure Table.

Remember that area is always a positive number, even if the z value is negative.

Situation 2 Find the area under the standard normal distribution curve in either tail.

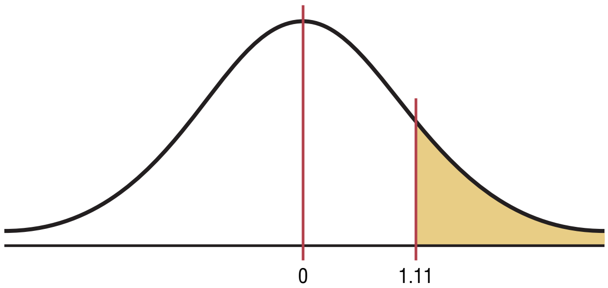
Example 6–4

Find the area to the right of $z = 1.11$.

Solution

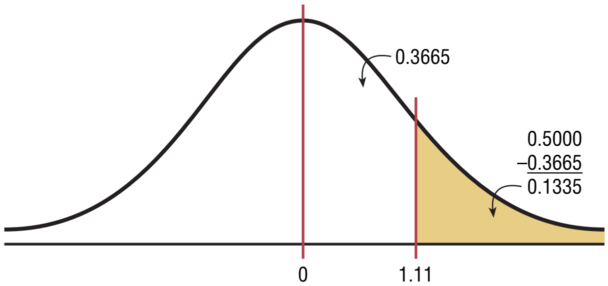
Draw the figure and represent the area as shown in Figure 6–11.

Figure 6–11
Area Under the Standard Normal Distribution Curve for Example 6–4



The required area is in the tail of the curve. Since Table E gives the area between $z = 0$ and $z = 1.11$, first find that area. Then subtract this value from 0.5000, since one-half of the area under the curve is to the right of $z = 0$. See Figure 6–12.

Figure 6–12
Finding the Area in the Tail of the Standard Normal Distribution Curve (Example 6–4)



The area between $z = 0$ and $z = 1.11$ is 0.3665, and the area to the right of $z = 1.11$ is 0.1335, or 13.35%, obtained by subtracting 0.3665 from 0.5000.

Example 6–5

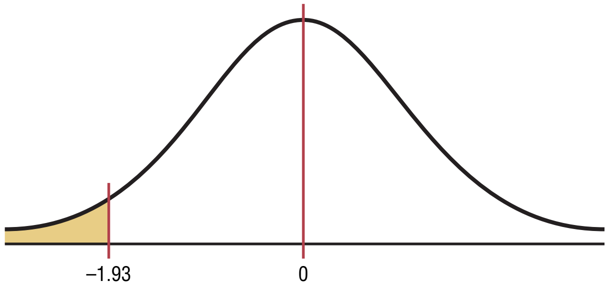
Find the area to the left of $z = -1.93$.

Solution

The desired area is shown in Figure 6–13.

Figure 6–13

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



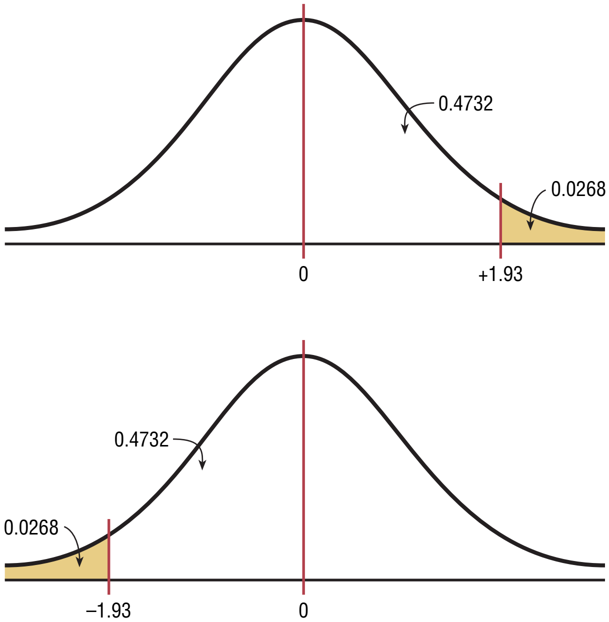
Again, Table E gives the area for positive z values. But from the symmetric property of the normal distribution, the area to the left of -1.93 is the same as the area to the right of $z = +1.93$, as shown in Figure 6–14.

Now find the area between 0 and $+1.93$ and subtract it from 0.5000, as shown:

$$\begin{array}{r} 0.5000 \\ -0.4732 \\ \hline 0.0268 \end{array}$$ or 2.68%

Figure 6–14

Comparison of Areas to the Right of +1.93 and to the Left of −1.93 (Example 6–5)



This procedure was summarized in block 2 of the Procedure Table.

Situation 3 Find the area under the standard normal distribution curve between any two z values on the same side of the mean.

Example 6–6

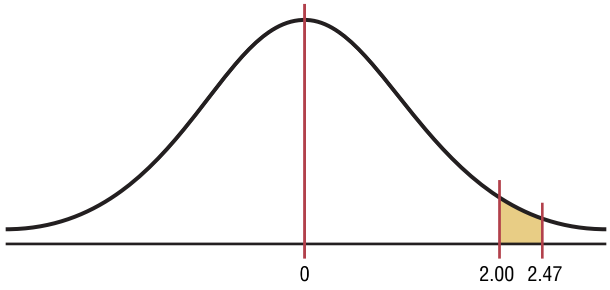
Find the area between $z = 2.00$ and $z = 2.47$.

Solution

The desired area is shown in Figure 6–15.

Figure 6–15

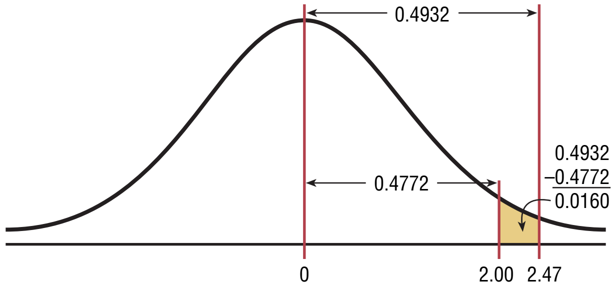
Area Under the Standard Normal Distribution Curve for Example 6–6



For this situation, look up the area from $z = 0$ to $z = 2.47$ and the area from $z = 0$ to $z = 2.00$. Then subtract the two areas, as shown in Figure 6–16.

Figure 6–16

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



The area between $z = 0$ and $z = 2.47$ is 0.4932. The area between $z = 0$ and $z = 2.00$ is 0.4772. Hence, the desired area is $0.4932 - 0.4772 = 0.0160$, or 1.60%. This procedure is summarized in block 3 of the Procedure Table.

Two things should be noted here. First, the *areas,* not the z values, are subtracted. Subtracting the z values will yield an incorrect answer. Second, the procedure in Example 6–6 is used when both z values are on the same side of the mean.

Example 6–7

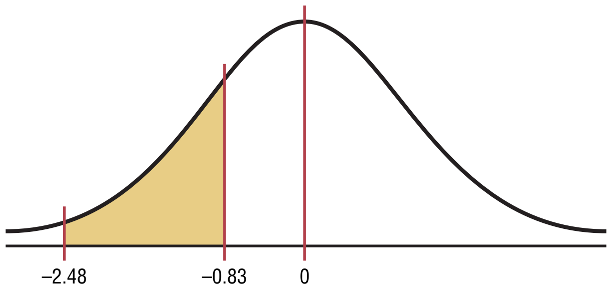
Find the area between $z = -2.48$ and $z = -0.83$.

Solution

The desired area is shown in Figure 6–17.

Figure 6–17

Area Under the Standard Normal Distribution Curve for Example 6–7



The area between $z = 0$ and $z = -2.48$ is 0.4934. The area between $z = 0$ and $z = -0.83$ is 0.2967. Subtracting yields $0.4934 - 0.2967 = 0.1967$, or 19.67%. This solution is summarized in block 3 of the Procedure Table.

Situation 4 Find the area under the standard normal distribution curve between any two z values on opposite sides of the mean.

Example 6–8

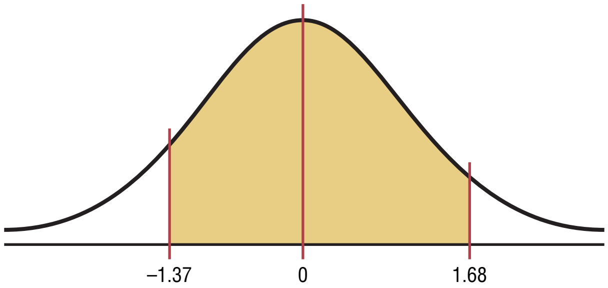
Find the area between $z = +1.68$ and $z = -1.37$.

Solution

The desired area is shown in Figure 6–18.

Figure 6–18

Area Under the Standard Normal Distribution Curve for Example 6–8



Now, since the two areas are on opposite sides of $z = 0$, one must find both areas and add them. The area between $z = 0$ and $z = 1.68$ is 0.4535. The area between $z = 0$ and $z = -1.37$ is 0.4147. Hence, the total area between $z = -1.37$ and $z = +1.68$ is $0.4535 + 0.4147 = 0.8682$, or 86.82%.

This type of problem is summarized in block 4 of the Procedure Table.

Situation 5 Find the area under the standard normal distribution curve to the left of any z value, where z is greater than the mean.

Example 6–9

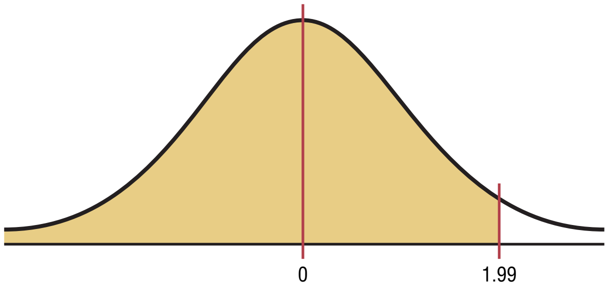
Find the area to the left of $z = 1.99$.

Solution

The desired area is shown in Figure 6–19.

Figure 6–19

Area Under the Standard Normal Distribution Curve for Example 6–9



Since Table E gives only the area between $z = 0$ and $z = 1.99$, one must add 0.5000 to the table area, since 0.5000 (one-half) of the total area lies to the left of $z = 0$. The area between $z = 0$ and $z = 1.99$ is 0.4767, and the total area is $0.4767 + 0.5000 = 0.9767$, or 97.67%.

This solution is summarized in block 5 of the Procedure Table.

The same procedure is used when the z value is to the left of the mean, as shown in Example 6–10.

Situation 6 Find the area under the standard normal distribution curve to the right of any z value, where z is less than the mean.

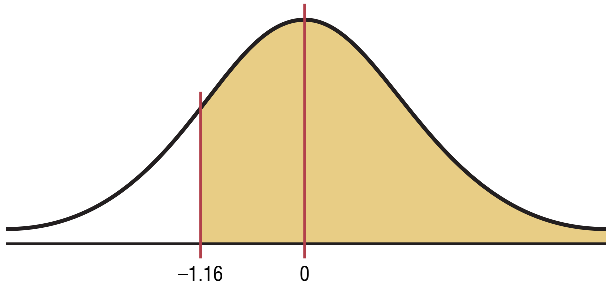
Example 6–10 Find the area to the right of $z = -1.16$.

Solution

The desired area is shown in Figure 6–20.

Figure 6–20

Area Under the Standard Normal Distribution Curve for Example 6–10



The area between $z = 0$ and $z = -1.16$ is 0.3770. Hence, the total area is $0.3770 + 0.5000 = 0.8770$, or 87.70%.

This type of problem is summarized in block 6 of the Procedure Table.

The final type of problem is that of finding the area in two tails. To solve it, find the area in each tail and add, as shown in Example 6–11.

Situation 7 Find the total area under the standard normal distribution curve in any two tails.

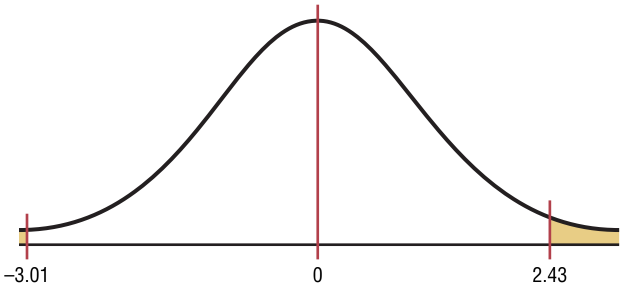
Example 6–11 Find the area to the right of $z = +2.43$ and to the left of $z = -3.01$.

Solution

The desired area is shown in Figure 6–21.

Figure 6–21

Area Under the Standard Normal Distribution Curve for Example 6–11



The area to the right of 2.43 is $0.5000 - 0.4925 = 0.0075$. The area to the left of $z = -3.01$ is $0.5000 - 0.4987 = 0.0013$. The total area, then, is $0.0075 + 0.0013 = 0.0088$, or 0.88%.

This solution is summarized in block 7 of the Procedure Table.

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 block 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 6–12

Find the probability for each.

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

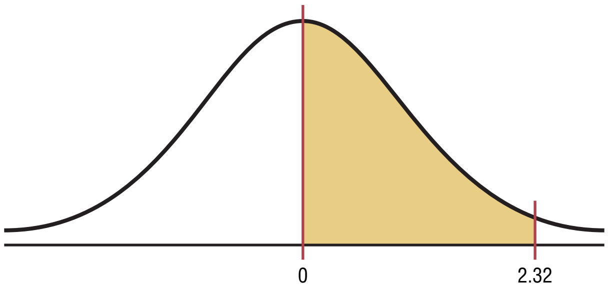
b. $P(z < 1.65)$

c. $P(z > 1.91)$

Solution

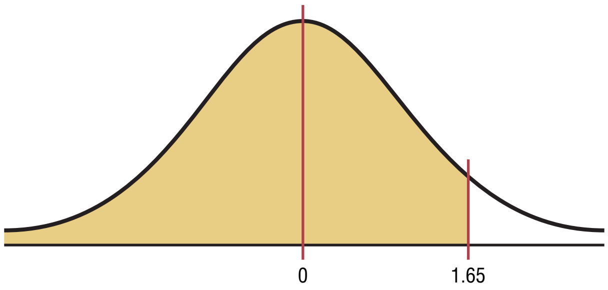
a. $P(0 < z < 2.32)$ means to find the area under the standard normal distribution curve between 0 and 2.32. Look up the area in Table E corresponding to $z = 2.32$. It is 0.4898, or 48.98%. The area is shown in Figure 6–22.

Figure 6–22
Area Under the Standard Normal Distribution Curve for Part *a* of Example 6–12



b. $P(z < 1.65)$ is represented in Figure 6–23.

Figure 6–23
Area Under the Standard Normal Distribution Curve for Part *b* of Example 6–12

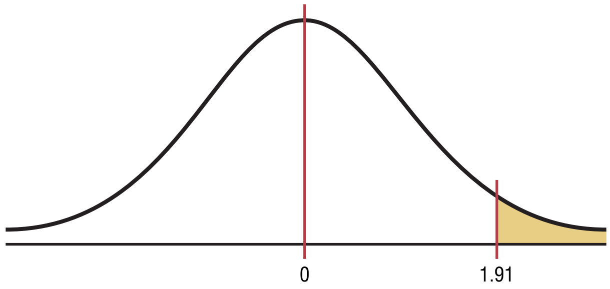


First, find the area between 0 and 1.65 in Table E. Then add it to 0.5000 to get 0.4505 + 0.5000 = 0.9505, or 95.05%.

c. $P(z > 1.91)$ is shown in Figure 6–24.

Figure 6–24

Area Under the Standard Normal Distribution Curve for Part *c* of Example 6–12



Since this area is a tail area, find the area between 0 and 1.91 and subtract it from 0.5000. Hence, 0.5000 − 0.4719 = 0.0281, or 2.81%.

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.

Example 6–13

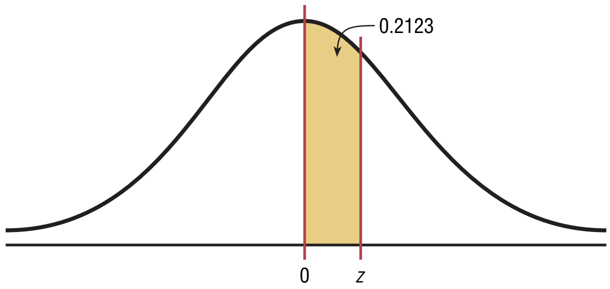
Find the *z* value such that the area under the standard normal distribution curve between 0 and the *z* value is 0.2123.

Solution

Draw the figure. The area is shown in Figure 6–25.

Figure 6–25

Area Under the Standard Normal Distribution Curve for Example 6–13



Next, find the area in Table E, as shown in Figure 6–26. Then read the correct *z* value in the left column as 0.5 and in the top row as 0.06, and add these two values to get 0.56.

Figure 6–26

Finding the *z* Value from Table E (Example 6–13)

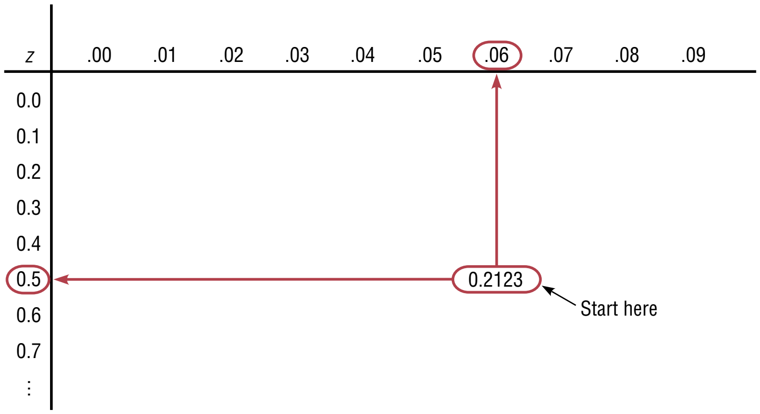
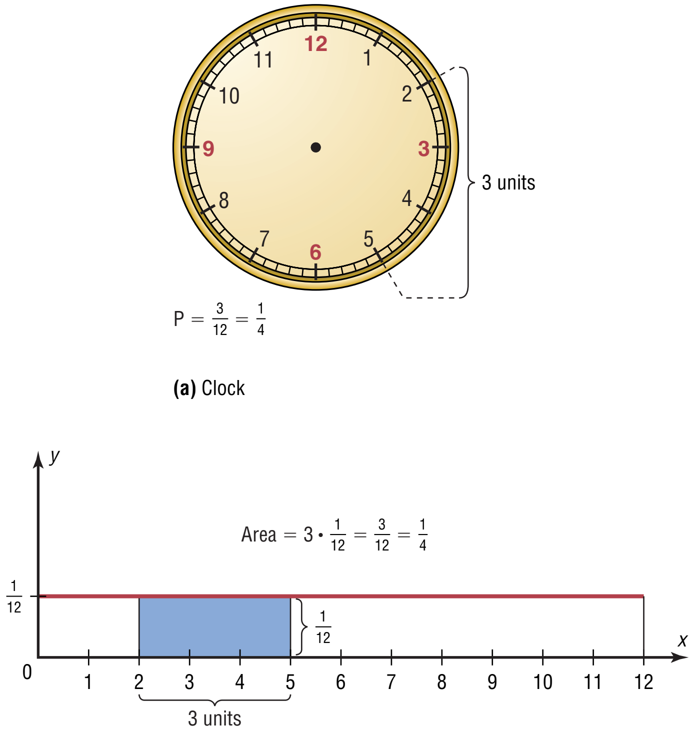


Figure 6–27

The Relationship Between Area and Probability



(a) Clock

(b) Rectangle

If the exact area cannot be found, use the closest value. For example, if one wanted to find the z value for an area 0.4241, the closest area is 0.4236, 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 6–27(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 would have 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 6–27(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 6–13.

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 6–4.

Applying the Concepts 6–3

Assessing Normality

Many times in statistics it is necessary to see if a distribution of data values is approximately normally distributed. There are special techniques that can be used. One technique is to draw a histogram for the data and see if it is approximately bell-shaped. (*Note:* It does not have to be exactly symmetric to be bell-shaped.)

The numbers of branches of the 50 top libraries are shown.

67	84	80	77	97	59	62	37	33	42
36	54	18	12	19	33	49	24	25	22
24	29	9	21	21	24	31	17	15	21
13	19	19	22	22	30	41	22	18	20
26	33	14	14	16	22	26	10	16	24

Source: *The World Almanac and Book of Facts.*

1. Construct a frequency distribution for the data.
2. Construct a histogram for the data.
3. Describe the shape of the histogram.
4. Based on your answer to question 3, do you feel that the distribution is approximately normal?

In addition to the histogram, distributions that are approximately normal have about 68% of the values fall within 1 standard deviation of the mean, about 95% of the data values fall within 2 standard deviations of the mean, and almost 100% of the data values fall within 3 standard deviations of the mean. (See Figure 6–5.)

5. Find the mean and standard deviation for the data.
6. What percent of the data values fall within 1 standard deviation of the mean?
7. What percent of the data values fall within 2 standard deviations of the mean?
8. What percent of data values fall within 3 standard deviations of the mean?
9. How do your answers to questions 6, 7, and 8 compare to 68, 95, and 100%, respectively?
10. Does your answer help support the conclusion you reached in question 4? Explain.

(More techniques for assessing normality are explained in Section 6–4.)
See page 344 for the answers.

Exercises 6–3

1. What are the characteristics of a normal distribution?

2. Why is the standard normal distribution important in statistical analysis?

3. What is the total area under the standard normal distribution curve?

4. What percentage of the area falls below the mean? Above the mean?

5. About what percentage of the area under the normal distribution curve falls within 1 standard deviation above and below the mean? 2 standard deviations? 3 standard deviations?

For Exercises 6 through 25, find the area under the standard normal distribution curve.

6. Between $z = 0$ and $z = 1.66$

7. Between $z = 0$ and $z = 0.75$

8. Between $z = 0$ and $z = -0.35$

9. Between $z = 0$ and $z = -2.07$

10. To the right of $z = 1.10$

11. To the right of $z = 0.23$

12. To the left of $z = -0.48$

13. To the left of $z = -1.43$

14. Between $z = 1.23$ and $z = 1.90$

15. Between $z = 0.79$ and $z = 1.28$

16. Between $z = -0.96$ and $z = -0.36$

17. Between $z = -1.56$ and $z = -1.83$

18. Between $z = 0.24$ and $z = -1.12$

19. Between $z = 2.47$ and $z = -1.03$

20. To the left of $z = 1.31$

21. To the left of $z = 2.11$

22. To the right of $z = -1.92$

23. To the right of $z = -0.15$

24. To the left of $z = -2.15$ and to the right of $z = 1.62$

25. To the right of $z = 1.92$ and to the left of $z = -0.44$

In Exercises 26 through 39, find probabilities for each, using the standard normal distribution.

26. $P(0 < z < 1.69)$

27. $P(0 < z < 0.67)$

28. $P(-1.23 < z < 0)$

29. $P(-1.57 < z < 0)$

30. $P(z > 1.16)$

31. $P(z > 2.83)$

32. $P(z < -1.77)$

33. $P(z < -1.21)$

34. $P(-0.05 < z < 1.10)$

35. $P(-2.46 < z < 1.74)$

36. $P(1.12 < z < 1.43)$

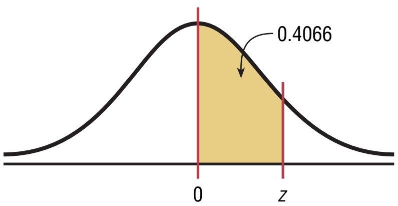
37. $P(1.46 < z < 2.97)$

38. $P(z > -1.39)$

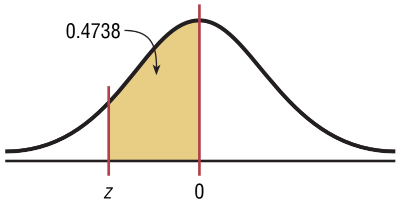
39. $P(z < 1.42)$

For Exercises 40 through 45, find the z value that corresponds to the given area.

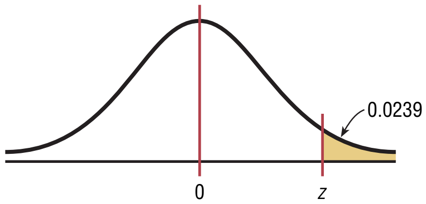
40.



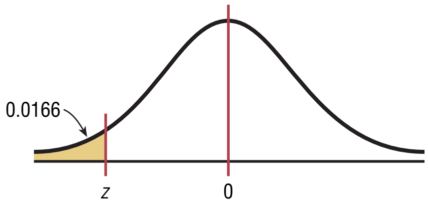
41.



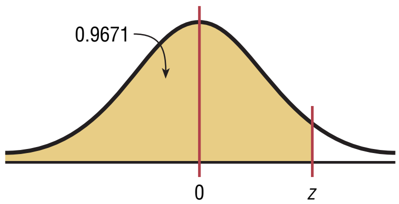
42.



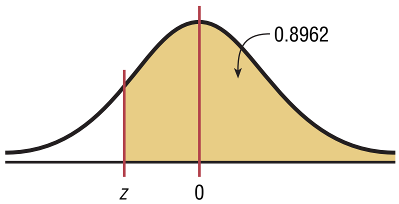
43.



44.



45.



46. Find the z value to the right of the mean so that

a. 53.98% of the area under the distribution curve lies to the left of it.

b. 71.90% of the area under the distribution curve lies to the left of it.

c. 96.78% of the area under the distribution curve lies to the left of it.

47. Find the z value to the left of the mean so that

a. 98.87% of the area under the distribution curve lies to the right of it.

b. 82.12% of the area under the distribution curve lies to the right of it.

c. 60.64% of the area under the distribution curve lies to the right of it.

48. Find two z values so that 44% of the middle area is bounded by them.

49. Find two z values, one positive and one negative, so that the areas in the two tails total the following values.

a. 5%
b. 10%
c. 1%

Extending the Concepts

50. In the standard normal distribution, find the values of z for the 75th, 80th, and 92nd percentiles.

51. Find $P(-1 < z < 1)$, $P(-2 < z < 2)$, and $P(-3 < z < 3)$. How do these values compare with the empirical rule?

52. Find z_0 such that $P(z > z_0) = 0.1234$.

53. Find z_0 such that $P(-1.2 < z < z_0) = 0.8671$.

54. Find z_0 such that $P(z_0 < z < 2.5) = 0.7672$.

55. Find z_0 such that the area between z_0 and $z = -0.5$ is 0.2345 (two answers).

56. Find z_0 such that $P(-z_0 < z < z_0) = 0.86$.

57. Find the equation for the standard normal distribution by substituting 0 for μ and 1 for σ in the equation

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

58. Graph by hand the standard normal distribution by using the formula derived in Exercise 57. Let $\pi \approx 3.14$ and $e \approx 2.718$. Use X values of $-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5$, and 2. (Use a calculator to compute the y values.)

Technology *Step by Step*

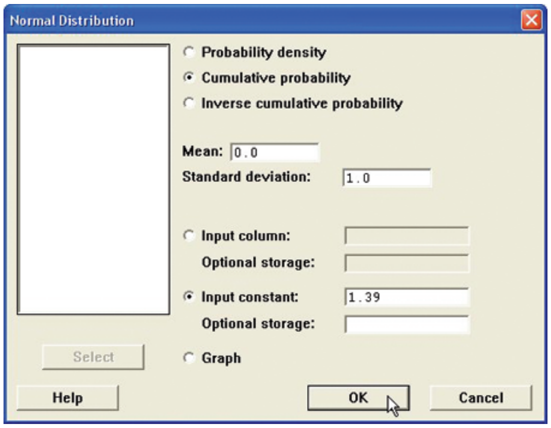
MINITAB Step by Step

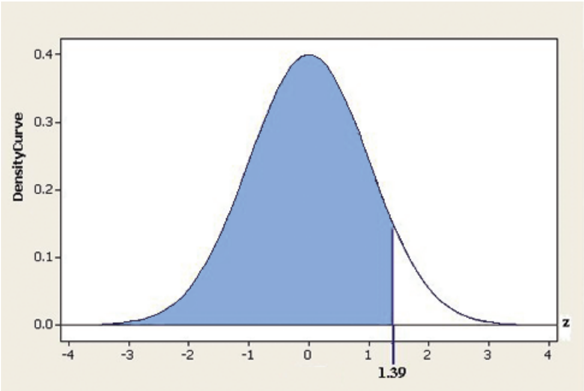
The Standard Normal Distribution

It is possible to determine the height of the density curve given a value of z, the cumulative area given a value of z, or a z value given a cumulative area. Examples are from Table E in Appendix C.

Find the Area to the Left of $z = 1.39$

1. Select **Calc>Probability Distributions>Normal.** There are three options.
2. Click the button for Cumulative probability. In the center section, the mean and standard deviation for the standard normal distribution are the defaults. The mean should be **0,** and the standard deviation should be **1.**
3. Click the button for Input Constant, then click inside the text box and type in **1.39.** Leave the storage box empty.
4. Click [OK].





Cumulative Distribution Function

Normal with mean = 0 and standard deviation = 1

x P(X <= x)

1.39 0.917736

The graph is not shown in the output.

The session window displays the result, 0.917736. If you choose the optional storage, type in a variable name such as **K1.** The result will be stored in the constant and will not be in the session window.

Find the Area to the Right of −2.06

1. Select **Calc>Probability Distributions>Normal.**
2. Click the button for Cumulative probability.
3. Click the button for Input Constant, then enter **−2.06** in the text box. Do not forget the minus sign.
4. Click in the text box for Optional storage and type **K1.**
5. Click [OK]. The area to the left of −2.06 is stored in K1 but not displayed in the session window.

   To determine the area to the right of the z value, subtract this constant from 1, then display the result.
6. Select **Calc>Calculator.**
   a) Type **K2** in the text box for Store result in:.
   b) Type in the expression **1 − K1,** then click [OK].
7. Select **Data>Display Data.** Drag the mouse over K1 and K2, then click [Select] and [OK].

   The results will be in the session window and stored in the constants.

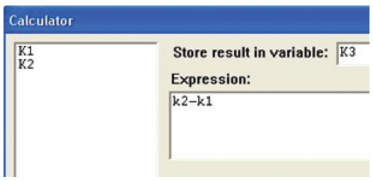
   Data Display

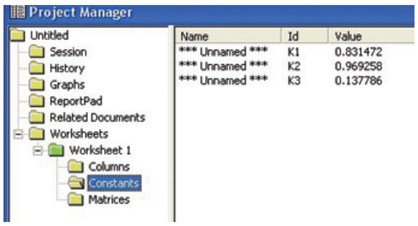
   K1 0.0196993

   K2 0.980301

8. To see the constants and other information about the worksheet, click the Project Manager icon. In the left pane click on the green worksheet icon, and then click the constants folder. You should see all constants and their values in the right pane of the Project Manager.
9. For the third example calculate the two probabilities and store them in K1 and K2.
10. Use the calculator to subtract K1 from K2 and store in K3.

   The calculator and project manager windows are shown.





Calculate a z Value Given the Cumulative Probability

Find the z value for a cumulative probability of 0.025.

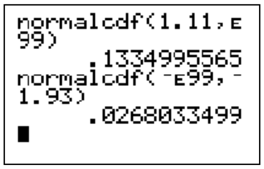
1. Select **Calc>Probability Distributions>Normal.**
2. Click the option for Inverse cumulative probability, then the option for Input constant.
3. In the text box type **.025,** the cumulative area, then click [OK].
4. In the dialog box, the z value will be returned, −1.960.

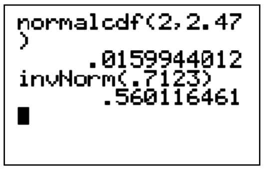
Inverse Cumulative Distribution Function
Normal with mean = 0 and standard deviation = 1

P (X <= x)	x
0.025	−1.95996

In the session window z is −1.95996.

TI-83 Plus or TI-84 Plus Step by Step





Standard Normal Random Variables

To find the probability for a standard normal random variable:
Press **2nd [DISTR],** then **2** for normalcdf(
The form is normalcdf(lower z score, upper z score).
Use E99 for ∞ (infinity) and −E99 for −∞ (negative infinity). Press **2nd [EE]** to get E.

Example: Area to the right of z = 1.11 (Example 6–4 from the text)
normalcdf(1.11,E99)

Example: Area to the left of z = −1.93 (Example 6–5 from the text)
normalcdf(−E99,−1.93)

Example: Area between z = 2.00 and z = 2.47 (Example 6–6 from the text)
normalcdf(2.00,2.47)

To find the percentile for a standard normal random variable:
Press **2nd [DISTR],** then **3** for the invNorm(
The form is invNorm(area to the left of z score)

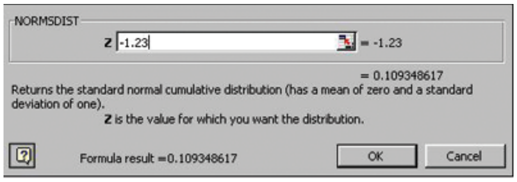
Example: Find the z score such that the area under the standard normal curve to the left of it is 0.7123 (Example 6–13 from the text)
invNorm(.7123)

Excel Step by Step

The Normal Distribution

To find area under the standard normal curve between two z values: $P(-1.23 < z < 2.54)$

1. Open a new worksheet and select a blank cell.
2. Click the f_x icon from the toolbar to call up the function list.
3. Select **NORMSDIST** from the Statistical function category.
4. Enter **−1.23** in the dialog box and click [OK]. This gives the area to the left of −1.23.
5. Select an adjacent blank cell, and repeat steps 1 through 4 for **2.54.**
6. To find the area between −1.23 and 2.54, select another blank cell and subtract the smaller area from the larger area.

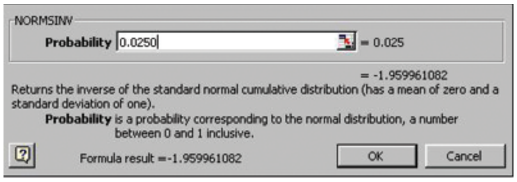


The area between the two values is the answer, 0.885109.

To find a z score corresponding to a cumulative area: $P(Z<=z) = 0.0250$

1. Click the f_x icon and select the Statistical function category.
2. Select the **NORMSINV** function and enter **0.0250.**
3. Click [OK].

The z score whose cumulative area is 0.0250 is the answer, -1.96.



6–4 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 Challenge on page 342. For all the problems presented in this chapter, one can assume that the variable is normally or approximately normally distributed.

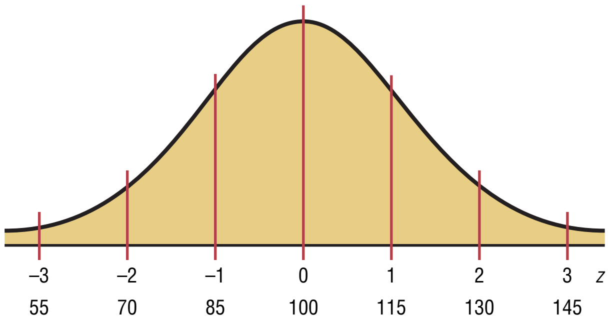
To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the formula

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

This is the same formula presented in Section 3–4. 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 6–28. (Recall that the z distribution has a mean of 0 and a standard deviation of 1.)

Figure 6–28
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 6–3.

Example 6–14

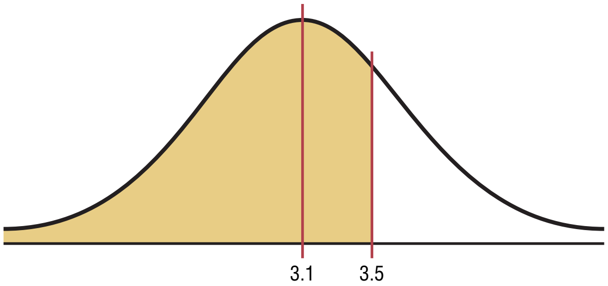
The mean number of hours an American worker spends on the computer is 3.1 hours per workday. Assume the standard deviation is 0.5 hour. Find the percentage of workers who spend less than 3.5 hours on the computer. Assume the variable is normally distributed.

Source: USA TODAY.

Solution

Step 1 Draw the figure and represent the area as shown in Figure 6–29.

Figure 6–29
Area Under a Normal Curve for Example 6–14

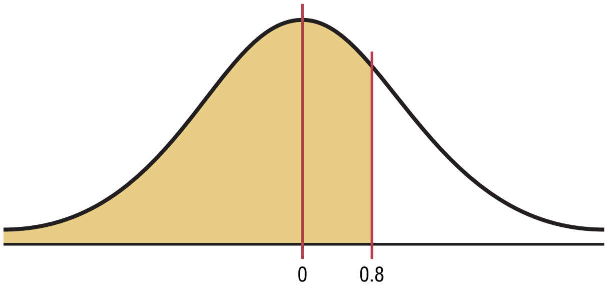


Step 2 Find the z value corresponding to 3.5.

$$z = \frac{X - \mu}{\sigma} = \frac{3.5 - 3.1}{0.5} = 0.80$$

Hence, 3.5 is 0.8 standard deviation above the mean of 3.1, as shown for the z distribution in Figure 6–30.

Figure 6–30
Area and z Values for Example 6–14



Step 3 Find the area by using Table E. The area between $z = 0$ and $z = 0.8$ is 0.2881. Since the area under the curve to the left of $z = 0.8$ is desired, add 0.5000 to 0.2881 $(0.5000 + 0.2881 = 0.7881)$.

Therefore, 78.81% of the workers spend less than 3.5 hours per workday on the computer.

Example 6–15

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.

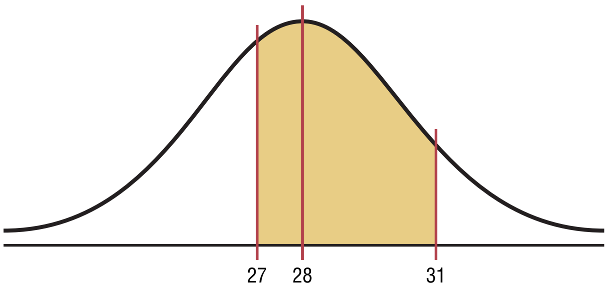
Assume the variable is approximately normally distributed.

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 6–31.

Figure 6–31
Area Under a Normal Curve for Part *a* of Example 6–15



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.

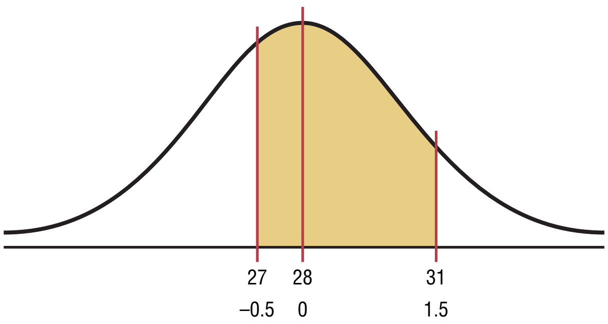
Step 2 Find the two z values.

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

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

Step 3 Find the appropriate area, using Table E. The area between $z = 0$ and $z = -0.5$ is 0.1915. The area between $z = 0$ and $z = 1.5$ is 0.4332. Add 0.1915 and 0.4332 ($0.1915 + 0.4332 = 0.6247$). Thus, the total area is 62.47%. See Figure 6–32.

Figure 6–32
Area and *z* Values for Part *a* of Example 6–15

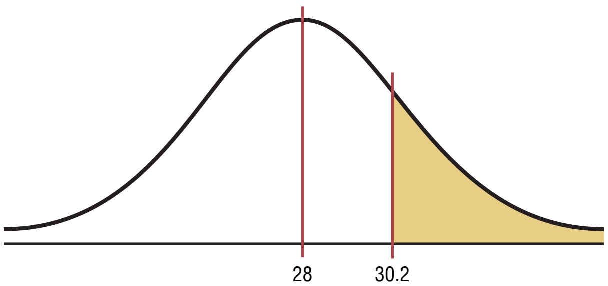


Hence, the probability that a randomly selected household generates between 27 and 31 pounds of newspapers per month is 62.47%.

Solution *b*

Step 1 Draw the figure and represent the area, as shown in Figure 6–33.

Figure 6–33
Area Under a Curve for Part *b* of Example 6–15



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 between $z = 0$ and $z = 1.1$ obtained from Table E is 0.3643. Since the desired area is in the right tail, subtract 0.3643 from 0.5000.

$0.5000 - 0.3643 = 0.1357$

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 6–16.

Example 6–16

The American Automobile Association reports that the average time it takes to respond to an emergency call is 25 minutes. Assume the variable is approximately normally distributed and the standard deviation is 4.5 minutes. If 80 calls are randomly selected, approximately how many will be responded to in less than 15 minutes?

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

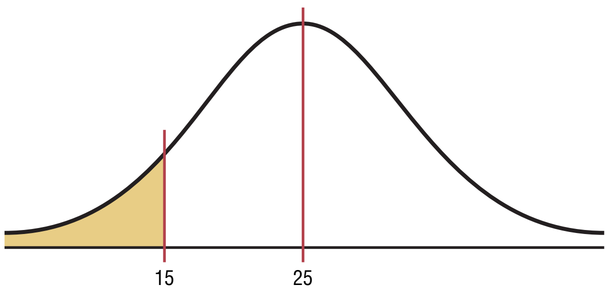
Solution

To solve the problem, find the area under a normal distribution curve to the left of 15.

Step 1 Draw a figure and represent the area as shown in Figure 6–34.

Figure 6–34

Area Under a Normal Curve for Example 6–16



Step 2 Find the z value for 15.

$$z = \frac{X - \mu}{\sigma} = \frac{15 - 25}{4.5} = -2.22$$

Step 3 Find the appropriate area. The area obtained from Table E is 0.4868, which corresponds to the area between $z = 0$ and $z = -2.22$. Use $+2.22$.

Step 4 Subtract 0.4868 from 0.5000 to get 0.0132.

Step 5 To find how many calls will be made in less than 15 minutes, multiply the sample size 80 by 0.0132 to get 1.056. Hence, 1.056, or approximately 1, call will be responded to in under 15 minutes.

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 1.056 calls.

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 6–17.

Example 6–17

Objective 5

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

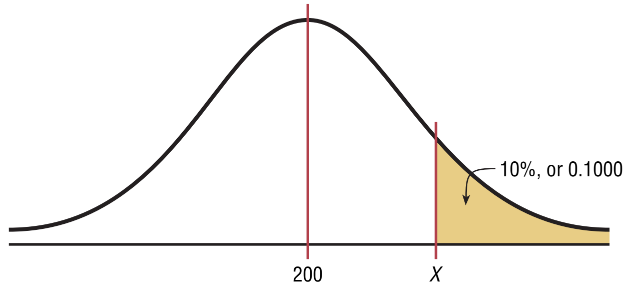
To qualify for a police academy, candidates must score in the top 10% on a general abilities test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume the test scores are normally distributed.

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 6–35.

Figure 6–35

Area Under a Normal Curve for Example 6–17



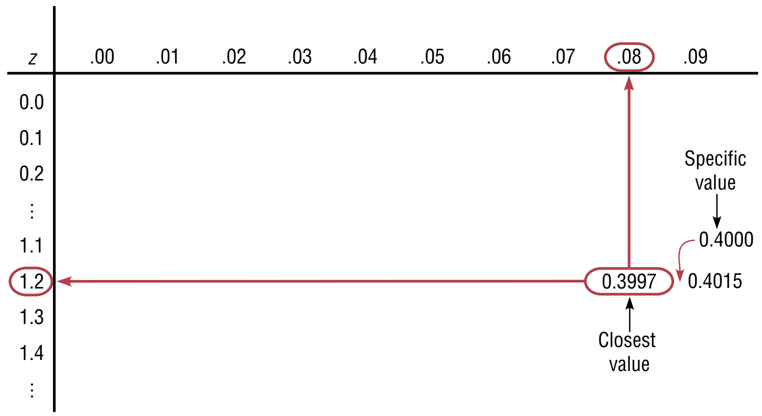
Work backward to solve this problem

Step 1 Subtract 0.1000 from 0.5000 to get the area under the normal distribution between 200 and *X*: 0.5000 − 0.1000 = 0.4000.

Step 2 Find the z value that corresponds to an area of 0.4000 by looking up 0.4000 in the area portion of Table E. If the specific value cannot be found, use the closest value—in this case 0.3997, as shown in Figure 6–36. 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 0.4500 falls halfway between 0.4495 and 0.4505. In this case use 1.65 rather than 1.64 for the z value.)

Figure 6–36

Finding the *z* Value from Table E (Example 6–17)



Interesting Fact

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

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

$$
\begin{aligned}
1.28 &= \frac{X - 200}{20} \\
(1.28)(20) + 200 &= X \\
25.60 + 200 &= X \\
225.60 &= X \\
226 &= X
\end{aligned}
$$

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

Instead of using the formula shown in step 3, one 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 one must find the value of X, the following formula can be used:

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

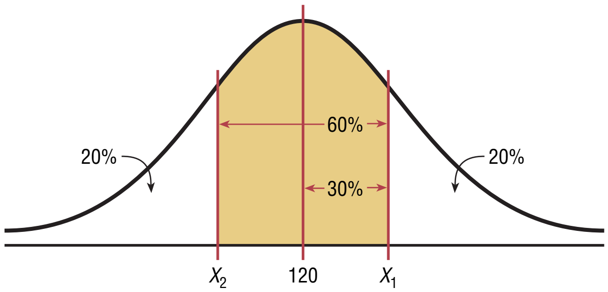
Example 6–18

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 6–37.

Figure 6–37 Area Under a Normal Curve for Example 6–18



Note that two values are needed, one above the mean and one below the mean. Find the value to the right of the mean first. The closest z value for an area of 0.3000 is 0.84. Substituting in the formula $X = z\sigma + \mu$, one gets

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

On the other side, $z = -0.84$; hence,

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

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.

Determining Normality

A normally shaped or bell-shaped distribution is only one of many shapes that a distribution can assume; however, it is very important since many statistical methods require that the distribution of values (shown in subsequent chapters) be normally or approximately normally shaped.

There are several ways statisticians check for normality. The easiest way is to draw a histogram for the data and check its shape. If the histogram is not approximately bell-shaped, then the data are not normally distributed.

Skewness can be checked by using Pearson's index PI of skewness. The formula is

$$\text{PI} = \frac{3(\bar{X} - \text{median})}{s}$$

If the index is greater than or equal to $+1$ or less than or equal to -1, it can be concluded that the data are significantly skewed.

In addition, the data should be checked for outliers by using the method shown in Chapter 3, page 143. Even one or two outliers can have a big effect on normality.

Examples 6–19 and 6–20 show how to check for normality.

Example 6–19

A survey of 18 high-technology firms showed the number of days' inventory they had on hand. Determine if the data are approximately normally distributed.

5	29	34	44	45	63	68	74	74
81	88	91	97	98	113	118	151	158

Source: USA TODAY.

Solution

Step 1 Construct a frequency distribution and draw a histogram for the data, as shown in Figure 6–38.

Class	Frequency
5–29	2
30–54	3
55–79	4
80–104	5
105–129	2
130–154	1
155–179	1

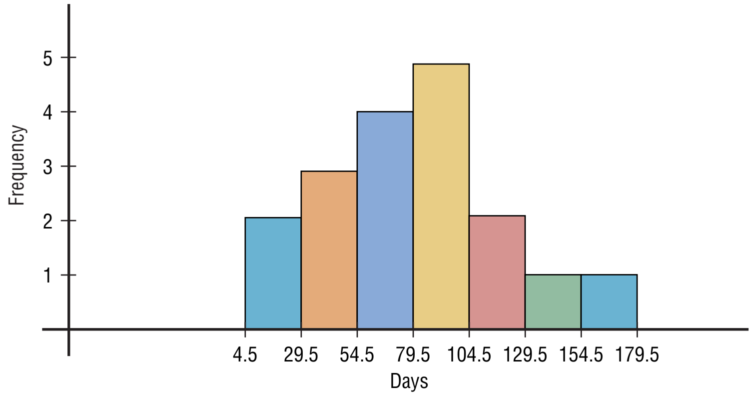


Figure 6–38
Histogram for Example 6–19

Since the histogram is approximately bell-shaped, one can say that the distribution is approximately normal.

Step 2 Check for skewness. For these data, $\overline{X} = 79.5$, median $= 77.5$, and $s = 40.5$. Using Pearson's index of skewness gives

$$\text{PI} = \frac{3(79.5 - 77.5)}{40.5}$$
$$= 0.148$$

In this case, the PI is not greater than $+1$ or less than -1, so it can be concluded that the distribution is not significantly skewed.

Step 3 Check for outliers. Recall that an outlier is a data value that lies more than 1.5 (IQR) units below Q_1 or 1.5 (IQR) units above Q_3. In this case, $Q_1 = 45$ and $Q_3 = 98$; hence, $\text{IQR} = Q_3 - Q_1 = 98 - 45 = 53$. An outlier would be a data value less than $45 - 1.5(53) = -34.5$ or a data value larger than $98 + 1.5(53) = 177.5$. In this case, there are no outliers.

Since the histogram is approximately bell-shaped, the data are not significantly skewed, and there are no outliers, it can be concluded that the distribution is approximately normally distributed.

Example 6–20

The data shown consist of the number of games played each year in the career of Baseball Hall of Famer Bill Mazeroski. Determine if the data are approximately normally distributed.

81	148	152	135	151	152
159	142	34	162	130	162
163	143	67	112	70	

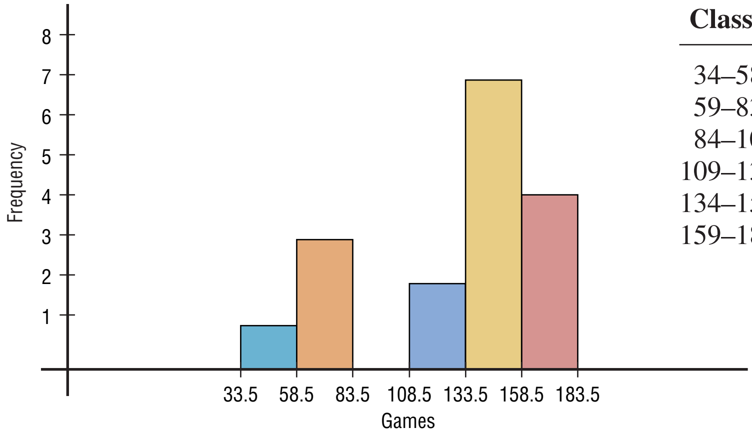
Source: Greensburg Tribune Review.

Solution

Step 1 Construct a frequency distribution and draw a histogram for the data. See Figure 6–39.

Figure 6–39

Histogram for Example 6–20



Class	Frequency
34–58	1
59–83	3
84–108	0
109–133	2
134–158	7
159–183	4

The histogram shows that the frequency distribution is somewhat negatively skewed.

Unusual Stats

The average amount of money stolen by a pickpocket each time is $128.

Step 2 Check for skewness; $\bar{X} = 127.24$, median $= 143$, and $s = 39.87$.

$$\begin{aligned} \text{PI} &= \frac{3(\bar{X} - \text{median})}{s} \\ &= \frac{3(127.24 - 143)}{39.87} \\ &= -1.19 \end{aligned}$$

Since the PI is less than -1, it can be concluded that the distribution is significantly skewed to the left.

Step 3 Check for outliers. In this case, $Q_1 = 96.5$ and $Q_3 = 155.5$. IQR $= Q_3 - Q_1 = 155.5 - 96.5 = 59$. Any value less than $96.5 - 1.5(59) = 8$ or above $155.5 + 1.5(59) = 244$ is considered an outlier. There are no outliers.

In summary, the distribution is somewhat negatively skewed.

Another method that is used to check normality is to draw a *normal quantile plot. Quantiles,* sometimes called *fractiles,* are values that separate the data set into approximately equal groups. Recall that quartiles separate the data set into four approximately equal groups, and deciles separate the data set into 10 approximately equal groups. A normal quantile plot consists of a graph of points using the data values for the x coordinates and the z values of the quantiles corresponding to the x values for the y coordinates. (*Note:* The calculations of the z values are somewhat complicated, and technology is usually used to draw the graph. The Technology Step by Step section shows how to draw a normal quantile plot.) If the points of the quantile plot do not lie in an approximately straight line, then normality can be rejected.

There are several other methods used to check for normality. A method using normal probability graph paper is shown in the Critical Thinking Challenge section at the end of this chapter, and the chi-square goodness-of-fit test is shown in Chapter 11. Two other tests sometimes used to check normality are the Kolmogorov-Smikirov test and the Lilliefors test. An explanation of these tests can be found in advanced textbooks.

Applying the Concepts 6–4

Smart People

Assume you are thinking about starting a Mensa chapter in your home town of Visiala, California, which has a population of about 10,000 people. You need to know how many people would qualify for Mensa, which requires an IQ of at least 130. You realize that IQ is normally distributed with a mean of 100 and a standard deviation of 15. Complete the following.

1. Find the approximate number of people in Visiala that are eligible for Mensa.
2. Is it reasonable to continue your quest for a Mensa chapter in Visiala?
3. How would you proceed to find out how many of the eligible people would actually join the new chapter? Be specific about your methods of gathering data.
4. What would be the minimum IQ score needed if you wanted to start an Ultra-Mensa club that included only the top 1% of IQ scores?

See page 344 for the answers.

Exercises 6–4

1. The average admission charge for a movie is \$5.39. If the distribution of admission charges is normal with a standard deviation of \$0.79, what is the probability that a randomly selected admission charge is less than \$3.00?

Source: *N.Y. Times Almanac.*

2. The average salary for first-year teachers is \$27,989. If the distribution is approximately normal with $\sigma = \$3250$, what is the probability that a randomly selected first-year teacher makes these salaries?

a. Between \$20,000 and \$30,000 a year

b. Less than \$20,000 a year

Source: *N.Y. Times Almanac.*

3. The average daily jail population in the United States is 618,319. If the distribution is normal and the standard deviation is 50,200, find the probability that on a randomly selected day the jail population is

a. Greater than 700,000.

b. Between 500,000 and 600,000.

Source: *N.Y. Times Almanac.*

4. The national average SAT score is 1019. If we assume a normal distribution with $\sigma = 90$, what is the 90th percentile score? What is the probability that a randomly selected score exceeds 1200?

Source: *N.Y. Times Almanac.*

5. The average number of calories in a 1.5-ounce chocolate bar is 225. Suppose that the distribution of calories is approximately normal with $\sigma = 10$. Find the probability that a randomly selected chocolate bar will have

a. Between 200 and 220 calories.

b. Less than 200 calories.

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

6. The average age of CEOs is 56 years. Assume the variable is normally distributed. If the standard deviation is 4 years, find the probability that the age of a randomly selected CEO will be in the following range.

a. Between 53 and 59 years old

b. Between 58 and 63 years old

c. Between 50 and 55 years old

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

7. The average salary for a Queens College full professor is \$85,900. If the average salaries are normally distributed with a standard deviation of \$11,000, find these probabilities.

a. The professor makes more than \$90,000.

b. The professor makes more than \$75,000.

Source: AAUP, *Chronicle of Higher Education.*

8. Full-time Ph.D. students receive an average of \$12,837 per year. If the average salaries are normally distributed with a standard deviation of \$1500, find these probabilities.

a. The student makes more than \$15,000.

b. The student makes between \$13,000 and \$14,000.

Source: U.S. Education Dept., *Chronicle of Higher Education.*

9. A survey found that people keep their microwave ovens an average of 3.2 years. The standard deviation is 0.56 year. If a person decides to buy a new microwave oven, find the probability that he or she has owned the old oven for the following amount of time. Assume the variable is normally distributed.

a. Less than 1.5 years

b. Between 2 and 3 years

c. More than 3.2 years

d. What percent of microwave ovens would be replaced if a warranty of 18 months were given?

10. The average commute to work (one way) is 25.5 minutes according to the 2000 Census. If we assume that commuting times are normally distributed with a standard deviation of 6.1 minutes, what is the probability that a randomly selected commuter spends more than 30 minutes a day commuting one way?

Source: *N.Y. Times Almanac.*

11. The average credit card debt for college seniors is \$3262. If the debt is normally distributed with a standard deviation of \$1100, find these probabilities.

a. That the senior owes at least \$1000

b. That the senior owes more than \$4000

c. That the senior owes between \$3000 and \$4000

Source: *USA TODAY.*

12. The average time a person spends at the Barefoot Landing Seaquarium is 96 minutes. The standard deviation is 17 minutes. Assume the variable is normally distributed. If a visitor is selected at random, find the probability that he or she will spend the following time at the seaquarium.

a. At least 120 minutes

b. At most 80 minutes

c. Suggest a time for a bus to return to pick up a group of tourists.

13. The average time for a mail carrier to cover his route is 380 minutes, and the standard deviation is 16 minutes. If one of these trips is selected at random, find the probability that the carrier will have the following route time. Assume the variable is normally distributed.

a. At least 350 minutes

b. At most 395 minutes

c. How might a mail carrier estimate a range for the time he or she will spend en route?

14. During October, the average temperature of Whitman Lake is 53.2° and the standard deviation is 2.3°. Assume the variable is normally distributed. For a randomly selected day in October, find the probability that the temperature will be as follows.

a. Above 54°

b. Below 60°

c. Between 49 and 55°

d. If the lake temperature were above 60°, would you call it very warm?

15. The average waiting time to be seated for dinner at a popular restaurant is 23.5 minutes, with a standard deviation of 3.6 minutes. Assume the variable is normally distributed. When a patron arrives at the restaurant for dinner, find the probability that the patron will have to wait the following time.

a. Between 15 and 22 minutes

b. Less than 18 minutes or more than 25 minutes

c. Is it likely that a person will be seated in less than 15 minutes?

16. A local medical research association proposes to sponsor a footrace. The average time it takes to run the course is 45.8 minutes with a standard deviation of 3.6 minutes. If the association decides to include only the top 25% of the racers, what should be the cutoff time in the tryout run? Assume the variable is normally distributed. Would a person who runs the course in 40 minutes qualify?

17. A marine sales dealer finds that the average price of a previously owned boat is \$6492. He decides to sell boats that will appeal to the middle 66% of the market in terms of price. Find the maximum and minimum prices of the boats the dealer will sell. The standard deviation is \$1025, and the variable is normally distributed. Would a boat priced at \$5550 be sold in this store?

18. The average charitable contribution itemized per income tax return in Pennsylvania is \$792. Suppose that the distribution of contributions is normal with a standard deviation of \$103. Find the limits for the middle 50% of contributions.

Source: IRS, *Statistics of Income Bulletin.*

19. A contractor decided to build homes that will include the middle 80% of the market. If the average size of homes built is 1810 square feet, find the maximum and minimum sizes of the homes the contractor should build. Assume that the standard deviation is 92 square feet and the variable is normally distributed.

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

20. If the average price of a new home is \$145,500, find the maximum and minimum prices of the houses that a contractor will build to include the middle 80% of the market. Assume that the standard deviation of prices is \$1500 and the variable is normally distributed.

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

21. The average price of a personal computer (PC) is \$949. If the computer prices are approximately normally distributed and $\sigma = \$100$, what is the probability that a randomly selected PC costs more than \$1200? The least expensive 10% of personal computers cost less than what amount?

Source: *N.Y. Times Almanac.*

22. To help students improve their reading, a school district decides to implement a reading program. It is to be administered to the bottom 5% of the students in the district, based on the scores on a reading achievement exam. If the average score for the students in the district is 122.6, find the cutoff score that will make a student eligible for the program. The standard deviation is 18. Assume the variable is normally distributed.

23. An automobile dealer finds that the average price of a previously owned vehicle is \$8256. He decides to sell cars that will appeal to the middle 60% of the market in terms of price. Find the maximum and minimum prices of the cars the dealer will sell. The standard deviation is \$1150, and the variable is normally distributed.

24. The average age of Amtrak passenger train cars is 19.4 years. If the distribution of ages is normal and 20% of the cars are older than 22.8 years, find the standard deviation.

Source: *N.Y. Times Almanac.*

25. The average length of a hospital stay is 5.9 days. If we assume a normal distribution and a standard deviation of 1.7 days, 15% of hospital stays are less than how many days? Twenty-five percent of hospital stays are longer than how many days?

Source: *N.Y. Times Almanac.*

26. A mandatory competency test for high school sophomores has a normal distribution with a mean of 400 and a standard deviation of 100.

a. The top 3% of students receive \$500. What is the minimum score you would need to receive this award?

b. The bottom 1.5% of students must go to summer school. What is the minimum score you would need to stay out of this group?

27. An advertising company plans to market a product to low-income families. A study states that for a particular area, the average income per family is \$24,596 and the standard deviation is \$6256. If the company plans to target the bottom 18% of the families based on income, find the cutoff income. Assume the variable is normally distributed.

28. If a one-person household spends an average of \$40 per week on groceries, find the maximum and minimum dollar amounts spent per week for the middle 50% of one-person households. Assume that the standard deviation is \$5 and the variable is normally distributed.

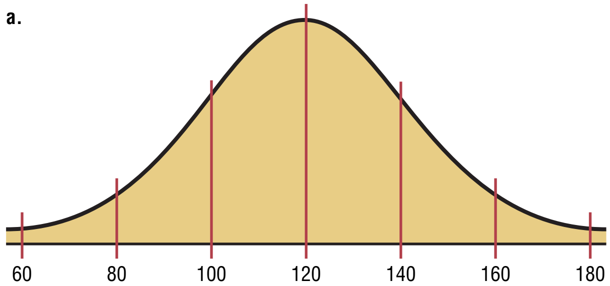
Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

29. The mean lifetime of a wristwatch is 25 months, with a standard deviation of 5 months. If the distribution is normal, for how many months should a guarantee be made if the manufacturer does not want to exchange more than 10% of the watches? Assume the variable is normally distributed.

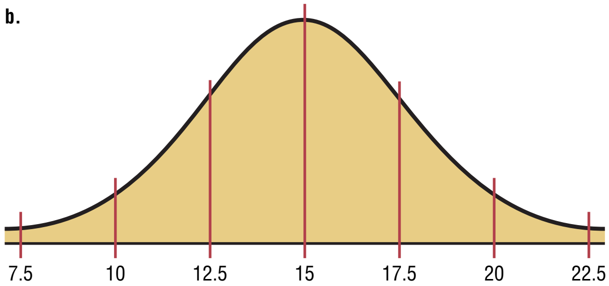
30. To qualify for security officers' training, recruits are tested for stress tolerance. The scores are normally distributed, with a mean of 62 and a standard deviation of 8. If only the top 15% of recruits are selected, find the cutoff score.

31. In the distributions shown, state the mean and standard deviation for each. *Hint:* See Figures 6–5 and 6–6. Also the vertical lines are 1 standard deviation apart.

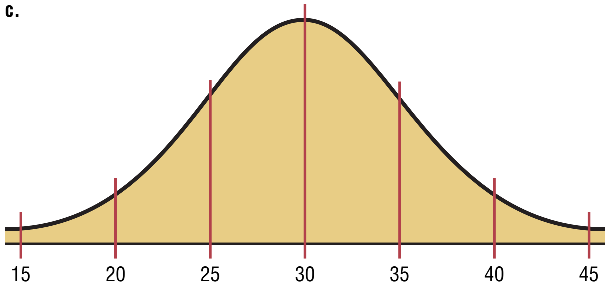
a.



b.



c.



32. Suppose that the mathematics SAT scores for high school seniors for a specific year have a mean of 456 and a standard deviation of 100 and are approximately normally distributed. If a subgroup of these high school seniors, those who are in the National Honor Society, is selected, would you expect the distribution of scores to have the same mean and standard deviation? Explain your answer.

33. Given a data set, how could you decide if the distribution of the data was approximately normal?

34. If a distribution of raw scores were plotted and then the scores were transformed to z scores, would the shape of the distribution change? Explain your answer.

35. In a normal distribution, find σ when $\mu = 110$ and 2.87% of the area lies to the right of 112.

36. In a normal distribution, find μ when σ is 6 and 3.75% of the area lies to the left of 85.

37. In a certain normal distribution, 1.25% of the area lies to the left of 42, and 1.25% of the area lies to the right of 48. Find μ and σ.

38. An instructor gives a 100-point examination in which the grades are normally distributed. The mean is 60 and the standard deviation is 10. If there are 5% A's and 5% F's, 15% B's and 15% D's, and 60% C's, find the scores that divide the distribution into those categories.

39. The data shown represent the number of outdoor drive-in movies in the United States for a 14-year period. Check for normality.

2084	1497	1014	910	899	870	837	859
848	826	815	750	637	737		

Source: National Association of Theater Owners.

40. The data shown represent the cigarette tax (in cents) for 30 randomly selected states. Check for normality.

3	58	5	65	17	48	52	75	21	76	58	36
100	111	34	41	23	44	33	50	13	18	7	12
20	24	66	28	28	31						

Source: Commerce Clearing House.

41. The data shown represent the box office total revenue (in millions of dollars) for a randomly selected sample of the top-grossing films in 2001. Check for normality.

294	241	130	144	113	70	97	94	91	202	74	79
71	67	67	56	180	199	165	114	60	56	53	51

Source: *USA TODAY.*

42. The data shown represent the number of runs made each year during Bill Mazeroski's career. Check for normality.

30	59	69	50	58	71	55	43	66	52	56	62
36	13	29	17	3							

Source: *Greensburg Tribune Review.*

Technology *Step by Step*

MINITAB
Step by Step

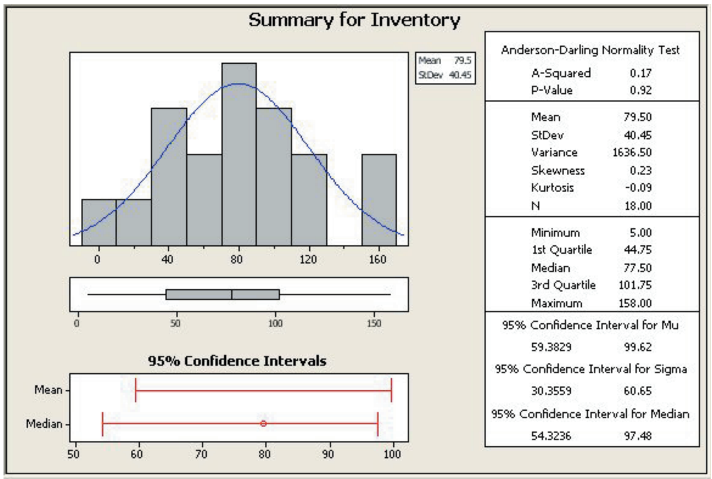
Determining Normality

There are several ways in which statisticians test a data set for normality. Four are shown here.

Construct a Histogram

Inspect the histogram for shape.

1. Enter the data for Example 6–19 in the first column of a new worksheet. Name the column Inventory.
2. Use **Stat>Basic Statistics>Graphical Summary** presented in Section 3–4 to create the histogram. Is it symmetric? Is there a single peak?



Check for Outliers

Inspect the boxplot for outliers. There are no outliers in this graph. Furthermore, the box is in the middle of the range, and the median is in the middle of the box. Most likely this is not a skewed distribution either.

Calculate Pearson's Index of Skewness

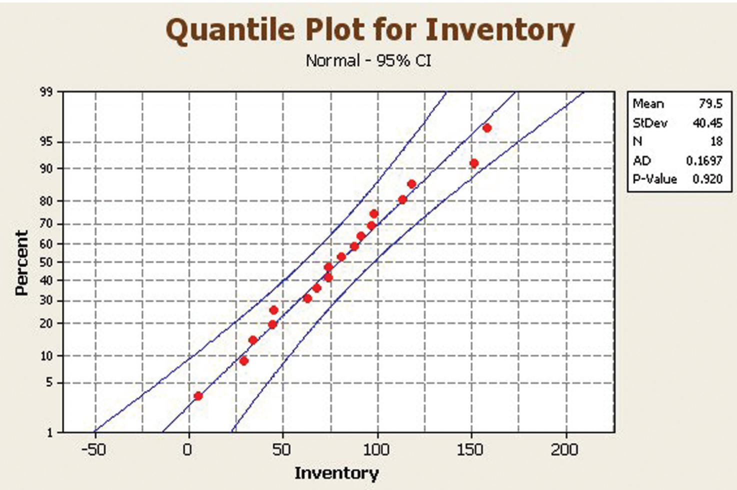
The measure of skewness in the graphical summary is not the same as Pearson's index. Use the calculator and the formula.

$$\text{PI} = \frac{3(\overline{X} - \text{median})}{s}$$

3. Select **Calc>Calculator,** then type **PI** in the text box for Store result in:.
4. Enter the expression: **3*(MEAN(C1)−MEDI(C1))/(STDEV(C1)).** Make sure you get all the parentheses in the right place!
5. Click [OK]. The result, 0.148318, will be stored in the first row of C2 named PI. Since it is smaller than +1, the distribution is not skewed.

Construct a Normal Probability Plot

6. Select **Graph>Probability Plot,** then Single and click [OK].
7. Double-click C1 Inventory to select the data to be graphed.
8. Click [Distribution] and make sure that Normal is selected. Click [OK].
9. Click [Labels] and enter the title for the graph: **Quantile Plot for Inventory.** You may also put **Your Name** in the subtitle.
10. Click [OK] twice. Inspect the graph to see if the graph of the points is linear.



These data are nearly normal.

What do you look for in the plot?

a) An "S curve" indicates a distribution that is too thick in the tails, a uniform distribution, for example.

b) Concave plots indicate a skewed distribution.

c) If one end has a point that is extremely high or low, there may be outliers.

This data set appears to be nearly normal by every one of the four criteria!

TI-83 Plus or TI-84 Plus
Step by Step

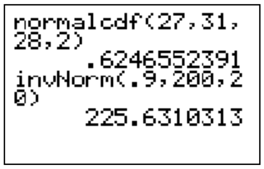
Normal Random Variables

To find the probability for a normal random variable:
Press **2nd [DISTR],** then **2** for normalcdf(
The form is normalcdf(lower *x* value, upper *x* value, μ, σ)
Use E99 for ∞ (infinity) and −E99 for −∞ (negative infinity). Press **2nd [EE]** to get E.

Example: Find the probability that *x* is between 27 and 31 when $\mu = 28$ and $\sigma = 2$ (Example 6–15*a* from the text).
normalcdf(27,31,28,2)

To find the percentile for a normal random variable:
Press **2nd [DISTR],** then **3** for invNorm(
The form is invNorm(area to the left of *x* value, μ, σ)

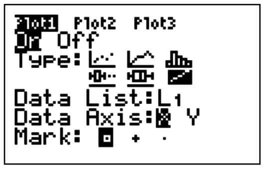
Example: Find the 90th percentile when $\mu = 200$ and $\sigma = 20$ (Example 6–17 from text).
invNorm(.9,200,20)

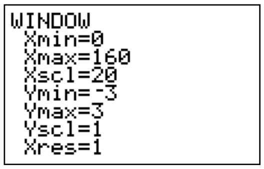


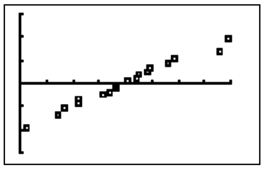
To construct a normal quantile plot:

1. Enter the data values into L_1.
2. Press **2nd [STAT PLOT]** to get the STAT PLOT menu.
3. Press **1** for Plot 1.
4. Turn on the plot by pressing **ENTER** while the cursor is flashing over ON.
5. Move the cursor to the normal quantile plot (6th graph).
6. Make sure L_1 is entered for the Data List and X is highlighted for the Data Axis.
7. Press **WINDOW** for the Window menu. Adjust Xmin and Xmax according to the data values. Adjust Ymin and Ymax as well, Ymin $= -3$ and Ymax $= 3$ usually work fine.
8. Press **GRAPH.**

Using the data from Example 6–19 gives







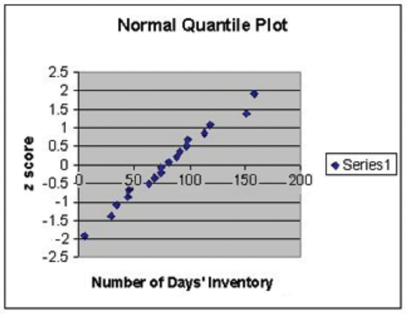
Since the points in the normal quantile plot lie close to a straight line, the distribution is approximately normal.

Excel
Step by Step

Normal Quantile Plot

Excel can be used to construct a normal quantile plot to examine if a set of data is approximately normally distributed.

1. Enter the data from Example 6–19 into column A of a new worksheet. The data should be sorted in ascending order.
2. Since the sample size is 18, each score represents $\frac{1}{18}$, or approximately 5.6%, of the sample. Each data point is assumed to subdivide the data into equal intervals. Each data value corresponds to the midpoint of the particular subinterval.
3. After all the data are entered and sorted in column A, select cell B1. From the function icon, select the NORMSINV command to find the z score corresponding to an area of $\frac{1}{18}$ of the total area under the normal curve. Enter **1/(2*18)** for the Probability.
4. Repeat the procedure from step 3 for each data value in column A. However, for each consecutive z score corresponding to a data value in column A, enter the next odd multiple of $\frac{1}{36}$ in the dialogue box. For example, in cell B2, enter the value **3/(2*18)** in the NORMSINV dialogue box. In cell B3, enter **5/(2*18).** Continue using this procedure to create z scores for each value in column A until all values have corresponding z scores.
5. Highlight the data from columns A and B, and select the Chart Wizard from the toolbar.
6. Select the scatter plot to graph the data from columns A and B as ordered pairs. Click Next.
7. Title and label axes as needed; click [OK].



The points appear to lie close to a straight line. Thus, we deduce that the data are approximately normally distributed.

6–5 The Central Limit Theorem

Objective 6

Use the central limit theorem to solve problems involving sample means for large samples.

In addition to knowing how individual data values vary about the mean for a population, statisticians are interested in knowing how the means of samples of the same size taken from the same population vary about the population mean.

Distribution of Sample Means

Suppose a researcher selects a sample of 30 adult males and finds the mean of the measure of the triglyceride levels for the sample subjects to be 187 milligrams/deciliter. Then suppose a second sample is selected, and the mean of that sample is found to be 192 milligrams/deciliter. Continue the process for 100 samples. What happens then is that the mean becomes a random variable, and the sample means 187, 192, 184, . . . , 196 constitute a *sampling distribution of sample means.*

A **sampling distribution of sample means** is a distribution using the means computed from all possible random samples of a specific size taken from a population.

If the samples are randomly selected with replacement, the sample means, for the most part, will be somewhat different from the population mean μ. These differences are caused by sampling error.

Sampling error is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

When all possible samples of a specific size are selected with replacement from a population, the distribution of the sample means for a variable has two important properties, which are explained next.

Properties of the Distribution of Sample Means

1. The mean of the sample means will be the same as the population mean.
2. The standard deviation of the sample means will be smaller than the standard deviation of the population, and it will be equal to the population standard deviation divided by the square root of the sample size.

The following example illustrates these two properties. Suppose a professor gave an 8-point quiz to a small class of four students. The results of the quiz were 2, 6, 4, and 8. For the sake of discussion, assume that the four students constitute the population. The mean of the population is

$$\mu = \frac{2 + 6 + 4 + 8}{4} = 5$$

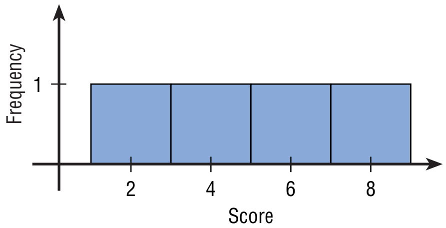
The standard deviation of the population is

$$\sigma = \sqrt{\frac{(2-5)^2 + (6-5)^2 + (4-5)^2 + (8-5)^2}{4}} = 2.236$$

The graph of the original distribution is shown in Figure 6–40. This is called a *uniform distribution.*

Figure 6–40

Distribution of Quiz Scores



Historical Notes

Two mathematicians who contributed to the development of the central limit theorem were Abraham DeMoivre (1667–1754) and Pierre Simon Laplace (1749–1827). DeMoivre was once jailed for his religious beliefs. After his release, DeMoivre made a living by consulting on the mathematics of gambling and insurance. He wrote two books, *Annuities Upon Lives* and *The Doctrine of Chance.*

Laplace held a government position under Napoleon and later under Louis XVIII. He once computed the probability of the sun rising to be 18,226,214/18,226,215.

Now, if all samples of size 2 are taken with replacement and the mean of each sample is found, the distribution is as shown.

Sample	Mean	Sample	Mean
2, 2	2	6, 2	4
2, 4	3	6, 4	5
2, 6	4	6, 6	6
2, 8	5	6, 8	7
4, 2	3	8, 2	5
4, 4	4	8, 4	6
4, 6	5	8, 6	7
4, 8	6	8, 8	8

A frequency distribution of sample means is as follows.

$\overline{X}$	f
2	1
3	2
4	3
5	4
6	3
7	2
8	1

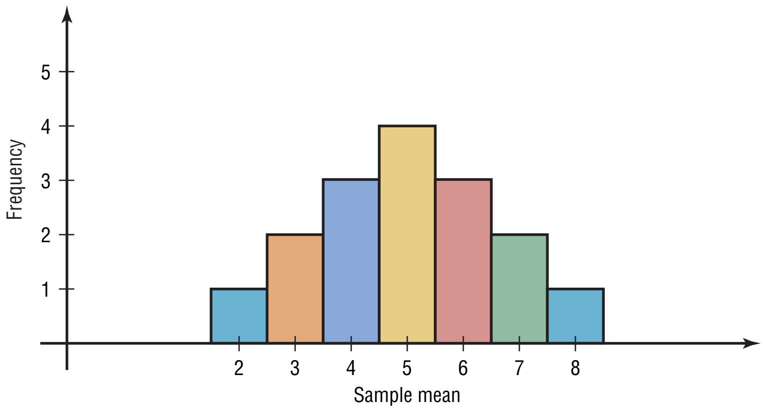
For the data from the example just discussed, Figure 6–41 shows the graph of the sample means. The histogram appears to be approximately normal.

The mean of the sample means, denoted by $\mu_{\overline{X}}$, is

$$\mu_{\overline{X}} = \frac{2 + 3 + \cdots + 8}{16} = \frac{80}{16} = 5$$

Figure 6–41

Distribution of Sample Means



which is the same as the population mean. Hence,

$$\mu_{\overline{X}} = \mu$$

The standard deviation of sample means, denoted by $\sigma_{\overline{X}}$, is

$$\sigma_{\overline{X}} = \sqrt{\frac{(2-5)^2 + (3-5)^2 + \cdots + (8-5)^2}{16}} = 1.581$$

which is the same as the population standard deviation, divided by $\sqrt{2}$:

$$\sigma_{\overline{X}} = \frac{2.236}{\sqrt{2}} = 1.581$$

Unusual Stats

Each year a person living in the United States consumes on average 1400 pounds of food.

(*Note:* Rounding rules were not used here in order to show that the answers coincide.)

In summary, if all possible samples of size n are taken with replacement from the same population, the mean of the sample means, denoted by $\mu_{\overline{X}}$, equals the population mean μ; and the standard deviation of the sample means, denoted by $\sigma_{\overline{X}}$, equals $\sigma/\sqrt{n}$. The standard deviation of the sample means is called the **standard error of the mean.** Hence,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

A third property of the sampling distribution of sample means pertains to the shape of the distribution and is explained by the **central limit theorem.**

The Central Limit Theorem

As the sample size n increases without limit, the shape of the distribution of the sample means taken with replacement from a population with mean μ and standard deviation σ will approach a normal distribution. As previously shown, this distribution will have a mean μ and a standard deviation $\sigma/\sqrt{n}$.

If the sample size is sufficiently large, the central limit theorem can be used to answer questions about sample means in the same manner that a normal distribution can be used to answer questions about individual values. The only difference is that a new formula must be used for the z values. It is

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Notice that $\overline{X}$ is the sample mean, and the denominator must be adjusted since means are being used instead of individual data values. The denominator is the standard deviation of the sample means.

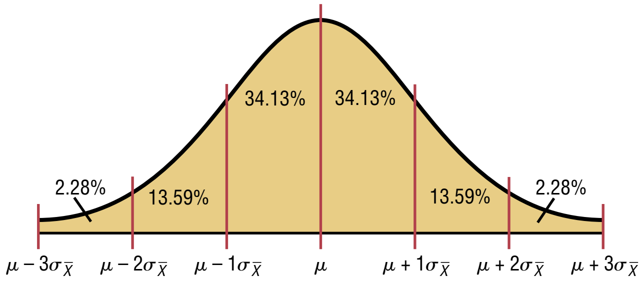
If a large number of samples of a given size are selected from a normally distributed population, or if a large number of samples of a given size that is greater than or equal to 30 are selected from a population that is not normally distributed, and the sample means are computed, then the distribution of sample means will look like the one shown in Figure 6–42. Their percentages indicate the areas of the regions.

It's important to remember two things when you use the central limit theorem:

1. When the original variable is normally distributed, the distribution of the sample means will be normally distributed, for any sample size n.
2. When the distribution of the original variable might not be normal, a sample size of 30 or more is needed to use a normal distribution to approximate the distribution of the sample means. The larger the sample, the better the approximation will be.

Figure 6–42

Distribution of Sample Means for Large Number of Samples



Examples 6–21 through 6–23 show how the standard normal distribution can be used to answer questions about sample means.

Example 6–21

A. C. Neilsen reported that children between the ages of 2 and 5 watch an average of 25 hours of television per week. Assume the variable is normally distributed and the standard deviation is 3 hours. If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the mean of the number of hours they watch television will be greater than 26.3 hours.

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

Solution

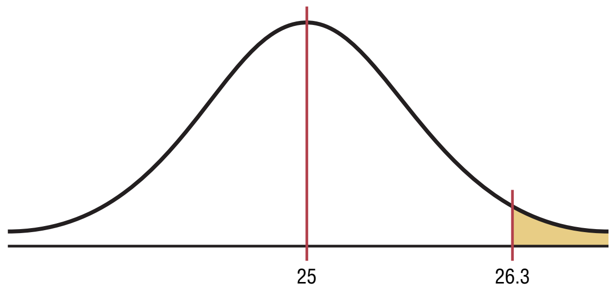
Since the variable is approximately normally distributed, the distribution of sample means will be approximately normal, with a mean of 25. The standard deviation of the sample means is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{20}} = 0.671$$

The distribution of the means is shown in Figure 6–43, with the appropriate area shaded.

Figure 6–43

Distribution of the Means for Example 6–21



The z value is

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{26.3 - 25}{3/\sqrt{20}} = \frac{1.3}{0.671} = 1.94$$

The area between 0 and 1.94 is 0.4738. Since the desired area is in the tail, subtract 0.4738 from 0.5000. Hence, 0.5000 − 0.4738 = 0.0262, or 2.62%.

One can conclude that the probability of obtaining a sample mean larger than 26.3 hours is 2.62% [i.e., $P(\overline{X} > 26.3) = 2.62\%$].

Example 6–22

The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected, find the probability that the mean of their age is between 90 and 100 months.

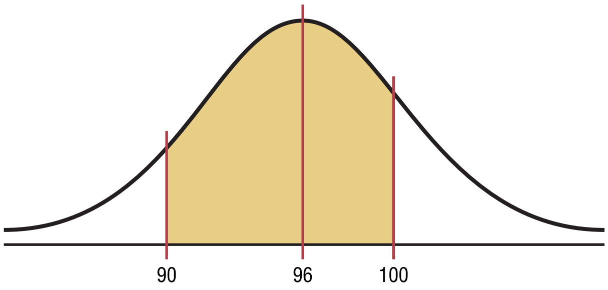
Source: Harper's Index.

Solution

Since the sample is 30 or larger, the normality assumption is not necessary. The desired area is shown in Figure 6–44.

Figure 6–44

Area Under a Normal Curve for Example 6–22



The two z values are

$$z_1 = \frac{90 - 96}{16/\sqrt{36}} = -2.25$$

$$z_2 = \frac{100 - 96}{16/\sqrt{36}} = 1.50$$

The two areas corresponding to the z values of -2.25 and 1.50, respectively, are 0.4878 and 0.4332. Since the z values are on opposite sides of the mean, find the probability by adding the areas: $0.4878 + 0.4332 = 0.921$, or 92.1%.

Hence, the probability of obtaining a sample mean between 90 and 100 months is 92.1%; that is, $P(90 < \overline{X} < 100) = 92.1\%$.

Students sometimes have difficulty deciding whether to use

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \quad \text{or} \quad z = \frac{X - \mu}{\sigma}$$

The formula

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

should be used to gain information about a sample mean, as shown in this section. The formula

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

is used to gain information about an individual data value obtained from the population. Notice that the first formula contains $\overline{X}$, the symbol for the sample mean, while the second formula contains X, the symbol for an individual data value. Example 6–23 illustrates the uses of the two formulas.

Example 6–23

The average number of pounds of meat that a person consumes a year is 218.4 pounds. Assume that the standard deviation is 25 pounds and the distribution is approximately normal.

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

a. Find the probability that a person selected at random consumes less than 224 pounds per year.

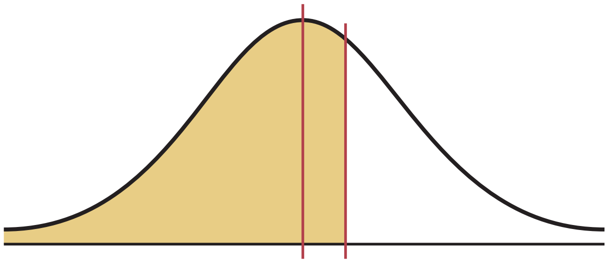
b. If a sample of 40 individuals is selected, find the probability that the mean of the sample will be less than 224 pounds per year.

Solution

a. Since the question asks about an individual person, the formula $z = (X - \mu)/\sigma$ is used. The distribution is shown in Figure 6–45.

Figure 6–45

Area Under a Normal Curve for Part *a* of Example 6–23



The z value is

$$z = \frac{X - \mu}{\sigma} = \frac{224 - 218.4}{25} = 0.22$$

The area between 0 and 0.22 is 0.0871; this area must be added to 0.5000 to get the total area to the left of $z = 0.22$.

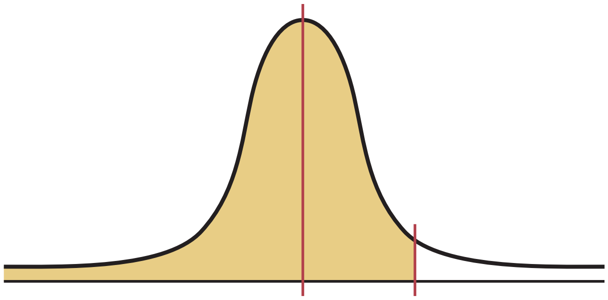
$$0.0871 + 0.5000 = 0.5871$$

Hence, the probability of selecting an individual who consumes less than 224 pounds of meat per year is 0.5871, or 58.71% [i.e., $P(X < 224) = 0.5871$].

b. Since the question concerns the mean of a sample with a size of 40, the formula $z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ is used. The area is shown in Figure 6–46.

Figure 6–46

Area Under a Normal Curve for Part *b* of Example 6–23



The z value is

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{224 - 218.4}{25/\sqrt{40}} = 1.42$$

The area between $z = 0$ and $z = 1.42$ is 0.4222; this value must be added to 0.5000 to get the total area.

$$0.4222 + 0.5000 = 0.9222$$

Hence, the probability that the mean of a sample of 40 individuals is less than 224 pounds per year is 0.9222, or 92.22%. That is, $P(\overline{X} < 224) = 0.9222$.

Comparing the two probabilities, one can see that the probability of selecting an individual who consumes less than 224 pounds of meat per year is 58.71%, but the probability of selecting a sample of 40 people with a mean consumption of meat that is less than 224 pounds per year is 92.22%. This rather large difference is due to the fact that the distribution of sample means is much less variable than the distribution of individual data values. (*Note:* An individual person is the equivalent of saying $n = 1$.)

Finite Population Correction Factor (Optional)

The formula for the standard error of the mean $\sigma/\sqrt{n}$ is accurate when the samples are drawn with replacement or are drawn without replacement from a very large or infinite population. Since sampling with replacement is for the most part unrealistic, a *correction factor* is necessary for computing the standard error of the mean for samples drawn without replacement from a finite population. Compute the correction factor by using the expression

$$\sqrt{\frac{N-n}{N-1}}$$

where N is the population size and n is the sample size.

This correction factor is necessary if relatively large samples are taken from a small population, because the sample mean will then more accurately estimate the population mean and there will be less error in the estimation. Therefore, the standard error of the mean must be multiplied by the correction factor to adjust for large samples taken from a small population. That is,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Finally, the formula for the z value becomes

$$z = \frac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}}$$

Interesting Fact

The bubonic plague killed more than 25 million people in Europe between 1347 and 1351.

When the population is large and the sample is small, the correction factor is generally not used, since it will be very close to 1.00.

The formulas and their uses are summarized in Table 6–1.

Table 6–1 Summary of Formulas and Their Uses

Formula	Use
1. $z = \dfrac{X - \mu}{\sigma}$	Used to gain information about an individual data value when the variable is normally distributed.
2. $z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$	Used to gain information when applying the central limit theorem about a sample mean when the variable is normally distributed or when the sample size is 30 or more.

Applying the Concepts 6–5

Central Limit Theorem

Twenty students from a statistics class each collected a random sample of times on how long it took students to get to class from their homes. All the sample sizes were 30. The resulting means are listed.

Student	Mean	Std. Dev.	Student	Mean	Std. Dev.
1	22	3.7	11	27	1.4
2	31	4.6	12	24	2.2
3	18	2.4	13	14	3.1
4	27	1.9	14	29	2.4
5	20	3.0	15	37	2.8
6	17	2.8	16	23	2.7
7	26	1.9	17	26	1.8
8	34	4.2	18	21.	2.0
9	23	2.6	19	30	2.2
10	29	2.1	20	29	2.8

1. The students noticed that everyone had different answers. If you randomly sample over and over from any population, with the same sample size, will the results ever be the same?
2. The students wondered whose results were right. How can they find out what the population mean and standard deviation are?
3. Input the means into the computer and check to see if the distribution is normal.
4. Check the mean and standard deviation of the means. How do these values compare to the students' individual scores?
5. Is the distribution of the means a sampling distribution?
6. Check the sampling error for students 3, 7, and 14.
7. Compare the standard deviation of the sample of the 20 means. Is that equal to the standard deviation from student 3 divided by the square of the sample size? How about for student 7, or 14?

See page 344 for the answers.

Exercises 6–5

1. If samples of a specific size are selected from a population and the means are computed, what is this distribution of means called?

2. Why do most of the sample means differ somewhat from the population mean? What is this difference called?

3. What is the mean of the sample means?

4. What is the standard deviation of the sample means called? What is the formula for this standard deviation?

5. What does the central limit theorem say about the shape of the distribution of sample means?

6. What formula is used to gain information about an individual data value when the variable is normally distributed?

7. What formula is used to gain information about a sample mean when the variable is normally distributed or when the sample size is 30 or more?

For Exercises 8 through 25, assume that the sample is taken from a large population and the correction factor can be ignored.

8. A survey found that the American family generates an average of 17.2 pounds of glass garbage each year. Assume the standard deviation of the distribution is 2.5 pounds. Find the probability that the mean of a sample of 55 families will be between 17 and 18 pounds.
Source: Michael D. Shook and Robert L. Shook, *The Book of Odds.*

9. The average yearly cost per household of owning a dog is \$186.80. Suppose that we randomly select 50 households that own a dog. What is the probability that the sample mean for these 50 households is less than \$175.00? Assume $\sigma = \$32$.

Source: *N.Y. Times Almanac.*

10. The average teacher's salary in New Jersey (ranked first among states) is \$52,174. Suppose that the distribution is normal with standard deviation equal to \$7500.

a. What is the probability that a randomly selected teacher makes less than \$50,000 a year?

b. If we sample 100 teachers' salaries, what is the probability that the sample mean is less than \$50,000?

Source: *N.Y. Times Almanac.*

11. The mean weight of 15-year-old males is 142 pounds, and the standard deviation is 12.3 pounds. If a sample of thirty-six 15-year-old males is selected, find the probability that the mean of the sample will be greater than 144.5 pounds. Assume the variable is normally distributed. Based on your answer, would you consider the group overweight?

12. The average teacher's salary in North Dakota is \$29,863. Assume a normal distribution with $\sigma = \$5100$.

a. What is the probability that a randomly selected teacher's salary is greater than \$40,000?

b. What is the probability that the mean for a sample of 80 teachers' salaries is greater than \$30,000?

Source: *N.Y. Times Almanac.*

13. The average price of a pound of sliced bacon is \$2.02. Assume the standard deviation is \$0.08. If a random sample of 40 one-pound packages is selected, find the probability that the mean of the sample will be less than \$2.00.

Source: *Statistical Abstract of the United States.*

14. The national average SAT score is 1019. Suppose that nothing is known about the shape of the distribution and that the standard deviation is 100. If a random sample of 200 scores were selected and the sample mean were calculated to be 1050, would you be surprised? Explain.

Source: *N.Y. Times Almanac.*

15. The average number of milligrams (mg) of sodium in a certain brand of low-salt microwave frozen dinners is 660 mg, and the standard deviation is 35 mg. Assume the variable is normally distributed.

a. If a single dinner is selected, find the probability that the sodium content will be more than 670 mg.

b. If a sample of 10 dinners is selected, find the probability that the mean of the sample will be larger than 670 mg.

c. Why is the probability for part *a* greater than that for part *b*?

16. The average age of chemical engineers is 37 years with a standard deviation of 4 years. If an engineering firm employs 25 chemical engineers, find the probability that the average age of the group is greater than 38.2 years old. If this is the case, would it be safe to assume that the engineers in this group are generally much older than average?

17. The *Old Farmer's Almanac* reports that the average person uses 123 gallons of water daily. If the standard deviation is 21 gallons, find the probability that the mean of a randomly selected sample of 15 people will be between 120 and 126 gallons. Assume the variable is normally distributed.

18. The average public elementary school has 458 students. Assume the standard deviation is 97. If a random sample of 36 public elementary schools is selected, find the probability that the number of students enrolled is between 450 and 465.

19. Procter & Gamble reported that an American family of four washes an average of 1 ton (2000 pounds) of clothes each year. If the standard deviation of the distribution is 187.5 pounds, find the probability that the mean of a randomly selected sample of 50 families of four will be between 1980 and 1990 pounds.

Source: *The Harper's Index Book.*

20. The average annual salary in Pennsylvania was \$24,393 in 1992. Assume that salaries were normally distributed for a certain group of wage earners, and the standard deviation of this group was \$4362.

a. Find the probability that a randomly selected individual earned less than \$26,000.

b. Find the probability that, for a randomly selected sample of 25 individuals, the mean salary was less than \$26,000.

c. Why is the probability for part *b* higher than the probability for part *a*?

Source: Associated Press.

21. The average time it takes a group of adults to complete a certain achievement test is 46.2 minutes. The standard deviation is 8 minutes. Assume the variable is normally distributed.

a. Find the probability that a randomly selected adult will complete the test in less than 43 minutes.

b. Find the probability that, if 50 randomly selected adults take the test, the mean time it takes the group to complete the test will be less than 43 minutes.

c. Does it seem reasonable that an adult would finish the test in less than 43 minutes? Explain.

d. Does it seem reasonable that the mean of the 50 adults could be less than 43 minutes?

22. Assume that the mean systolic blood pressure of normal adults is 120 millimeters of mercury (mm Hg) and the standard deviation is 5.6. Assume the variable is normally distributed.

a. If an individual is selected, find the probability that the individual's pressure will be between 120 and 121.8 mm Hg.

b. If a sample of 30 adults is randomly selected, find the probability that the sample mean will be between 120 and 121.8 mm Hg.

c. Why is the answer to part *a* so much smaller than the answer to part *b*?

23. The average cholesterol content of a certain brand of eggs is 215 milligrams, and the standard deviation is 15 milligrams. Assume the variable is normally distributed.

a. If a single egg is selected, find the probability that the cholesterol content will be greater than 220 milligrams.

b. If a sample of 25 eggs is selected, find the probability that the mean of the sample will be larger than 220 milligrams.

Source: *Living Fit.*

24. At a large publishing company, the mean age of proofreaders is 36.2 years, and the standard deviation is 3.7 years. Assume the variable is normally distributed.

a. If a proofreader from the company is randomly selected, find the probability that his or her age will be between 36 and 37.5 years.

b. If a random sample of 15 proofreaders is selected, find the probability that the mean age of the proofreaders in the sample will be between 36 and 37.5 years.

25. In the United States, one farmworker supplied agricultural products for an average of 106 people. Assume the standard deviation is 16.1. If 35 farmworkers are selected, find the probability that the mean number of people supplied is between 100 and 110.

Extending the Concepts

For Exercises 26 and 27, check to see whether the correction factor should be used. If so, be sure to include it in the calculations.

26. In a study of the life expectancy of 500 people in a certain geographic region, the mean age at death was 72.0 years, and the standard deviation was 5.3 years. If a sample of 50 people from this region is selected, find the probability that the mean life expectancy will be less than 70 years.

27. A study of 800 homeowners in a certain area showed that the average value of the homes was $82,000, and the standard deviation was $5000. If 50 homes are for sale, find the probability that the mean of the values of these homes is greater than $83,500.

28. The average breaking strength of a certain brand of steel cable is 2000 pounds, with a standard deviation of 100 pounds. A sample of 20 cables is selected and tested. Find the sample mean that will cut off the upper 95% of all samples of size 20 taken from the population. Assume the variable is normally distributed.

29. The standard deviation of a variable is 15. If a sample of 100 individuals is selected, compute the standard error of the mean. What size sample is necessary to double the standard error of the mean?

30. In Exercise 29, what size sample is needed to cut the standard error of the mean in half?

6–6 The Normal Approximation to the Binomial Distribution

A normal distribution is often used to solve problems that involve the binomial distribution since, when n is large (say, 100), the calculations are too difficult to do by hand using the binomial distribution. Recall from Chapter 5 that a binomial distribution has the following characteristics:

1. There must be a fixed number of trials.
2. The outcome of each trial must be independent.

3. Each experiment can have only two outcomes or outcomes that can be reduced to two outcomes.
4. The probability of a success must remain the same for each trial.

Also, recall that a binomial distribution is determined by n (the number of trials) and p (the probability of a success). When p is approximately 0.5, and as n increases, the shape of the binomial distribution becomes similar to that of a normal distribution. The larger n is and the closer p is to 0.5, the more similar the shape of the binomial distribution is to that of a normal distribution.

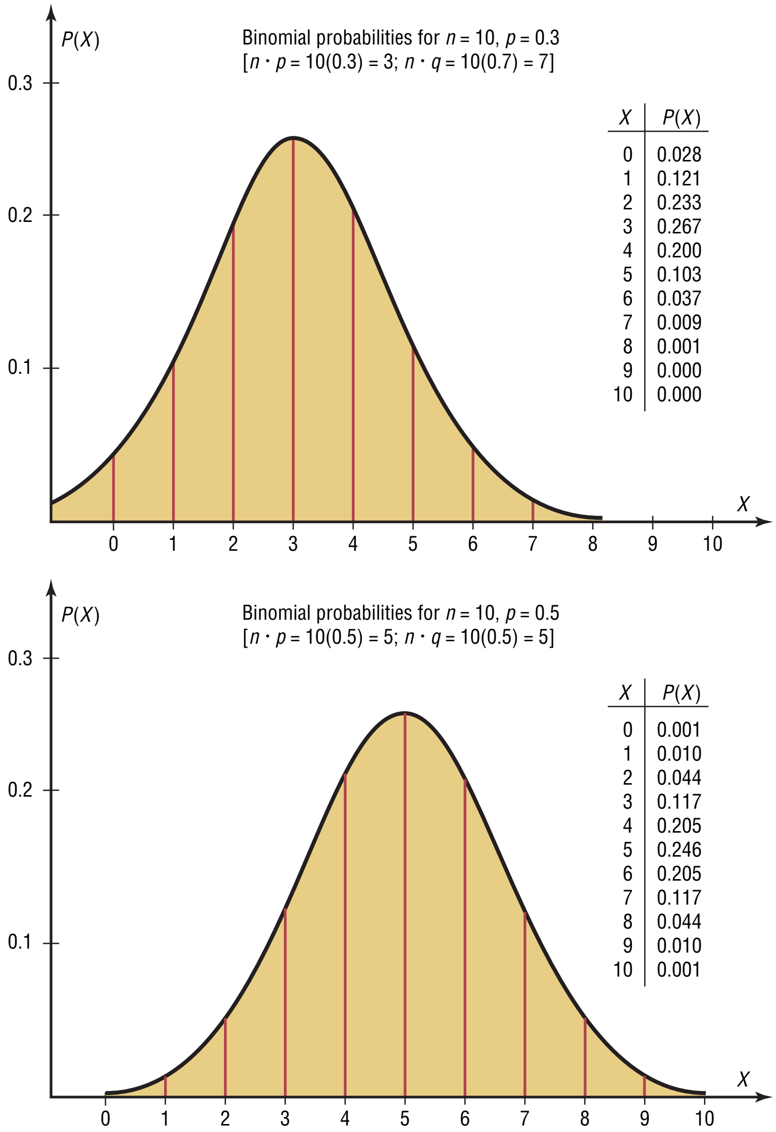
Objective 7

Use the normal approximation to compute probabilities for a binomial variable.

But when p is close to 0 or 1 and n is relatively small, a normal approximation is inaccurate. As a rule of thumb, statisticians generally agree that a normal approximation should be used only when $n \cdot p$ and $n \cdot q$ are both greater than or equal to 5. (*Note:* $q = 1 - p$.) For example, if p is 0.3 and n is 10, then $np = (10)(0.3) = 3$, and a normal distribution should not be used as an approximation. On the other hand, if $p = 0.5$ and $n = 10$, then $np = (10)(0.5) = 5$ and $nq = (10)(0.5) = 5$, and a normal distribution can be used as an approximation. See Figure 6–47.

Figure 6–47

Comparison of the Binomial Distribution and a Normal Distribution



In addition to the previous condition of $np \geq 5$ and $nq \geq 5$, a correction for continuity may be used in the normal approximation.

A **correction for continuity** is a correction employed when a continuous distribution is used to approximate a discrete distribution.

The continuity correction means that for any specific value of X, say 8, the boundaries of X in the binomial distribution (in this case, 7.5 to 8.5) must be used. (See Section 1–3.) Hence, when one employs a normal distribution to approximate the binomial, the boundaries of any specific value X must be used as they are shown in the binomial distribution. For example, for $P(X = 8)$, the correction is $P(7.5 < X < 8.5)$. For $P(X \leq 7)$, the correction is $P(X < 7.5)$. For $P(X \geq 3)$, the correction is $P(X > 2.5)$.

Students sometimes have difficulty deciding whether to add 0.5 or subtract 0.5 from the data value for the correction factor. Table 6–2 summarizes the different situations.

Table 6–2 Summary of the Normal Approximation to the Binomial Distribution

Binomial	Normal
When finding:	Use:
1. $P(X = a)$	$P(a - 0.5 < X < a + 0.5)$
2. $P(X \geq a)$	$P(X > a - 0.5)$
3. $P(X > a)$	$P(X > a + 0.5)$
4. $P(X \leq a)$	$P(X < a + 0.5)$
5. $P(X < a)$	$P(X < a - 0.5)$

For all cases, $\mu = n \cdot p$, $\sigma = \sqrt{n \cdot p \cdot q}$, $n \cdot p \geq 5$, and $n \cdot q \geq 5$.

Interesting Fact

Of the 12 months, August ranks first in the number of births for Americans.

The formulas for the mean and standard deviation for the binomial distribution are necessary for calculations. They are

$$\mu = n \cdot p \qquad \text{and} \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

The steps for using the normal distribution to approximate the binomial distribution are shown in this Procedure Table.

Procedure Table

Procedure for the Normal Approximation to the Binomial Distribution

Step 1 Check to see whether the normal approximation can be used.

Step 2 Find the mean μ and the standard deviation σ.

Step 3 Write the problem in probability notation, using X.

Step 4 Rewrite the problem by using the continuity correction factor, and show the corresponding area under the normal distribution.

Step 5 Find the corresponding z values.

Step 6 Find the solution.

Example 6–24

A magazine reported that 6% of American drivers read the newspaper while driving. If 300 drivers are selected at random, find the probability that exactly 25 say they read the newspaper while driving.

Source: USA Snapshot, *USA TODAY.*

Solution

Here, $p = 0.06$, $q = 0.94$, and $n = 300$.

Step 1 Check to see whether a normal approximation can be used.

$np = (300)(0.06) = 18 \qquad nq = (300)(0.94) = 282$

Since $np \geq 5$ and $nq \geq 5$, the normal distribution can be used.

Step 2 Find the mean and standard deviation.

$\mu = np = (300)(0.06) = 18$

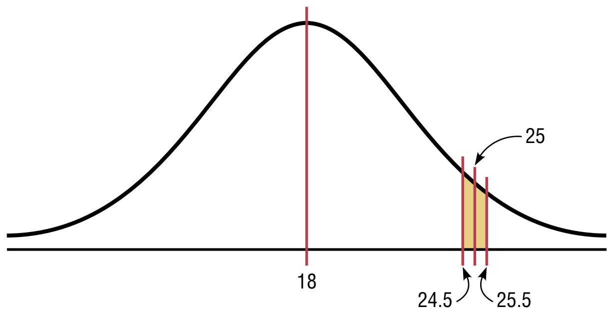
$\sigma = \sqrt{npq} = \sqrt{(300)(0.06)(0.94)} = \sqrt{16.92} = 4.11$

Step 3 Write the problem in probability notation: $P(X = 25)$.

Step 4 Rewrite the problem by using the continuity correction factor. See approximation number 1 in Table 6–2: $P(25 - 0.5 < X < 25 + 0.5) = P(24.5 < X < 25.5)$. Show the corresponding area under the normal distribution curve. See Figure 6–48.

Figure 6–48

Area Under a Normal Curve and *X* Values for Example 6–24



Step 5 Find the corresponding z values. Since 25 represents any value between 24.5 and 25.5, find both z values.

$$z_1 = \frac{25.5 - 18}{4.11} = 1.82 \qquad z_2 = \frac{24.5 - 18}{4.11} = 1.58$$

Step 6 Find the solution. Find the corresponding areas in the table: The area for $z = 1.82$ is 0.4656, and the area for $z = 1.58$ is 0.4429. Subtract the areas to get the approximate value: $0.4656 - 0.4429 = 0.0227$, or 2.27%.

Hence, the probability that exactly 25 people read the newspaper while driving is 2.27%.

Example 6–25

Of the members of a bowling league, 10% are widowed. If 200 bowling league members are selected at random, find the probability that 10 or more will be widowed.

Solution

Here, $p = 0.10$, $q = 0.90$, and $n = 200$.

Step 1 Since $np = (200)(0.10) = 20$ and $nq = (200)(0.90) = 180$, the normal approximation can be used.

Step 2 $\mu = np = (200)(0.10) = 20$

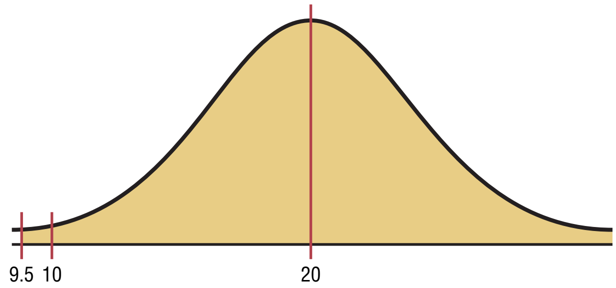
$$\sigma = \sqrt{npq} = \sqrt{(200)(0.10)(0.90)} = \sqrt{18} = 4.24$$

Step 3 $P(X \geq 10)$.

Step 4 See approximation number 2 in Table 6–2: $P(X > 10 - 0.5) = P(X > 9.5)$. The desired area is shown in Figure 6–49.

Figure 6–49

Area Under a Normal Curve and *X* Value for Example 6–25



Step 5 Since the problem is to find the probability of 10 or more positive responses, a normal distribution graph is as shown in Figure 6–49. Hence, the area between 9.5 and 20 must be added to 0.5000 to get the correct approximation.

The z value is

$$z = \frac{9.5 - 20}{4.24} = -2.48$$

Step 6 The area between 9.5 and 20 is 0.4934. Thus, the probability of getting 10 or more responses is $0.4934 + 0.5000 = 0.9934$, or 99.34%.

It can be concluded, then, that the probability of 10 or more widowed people in a random sample of 200 bowling league members is 99.34%.

Example 6–26

If a baseball player's batting average is 0.320 (32%), find the probability that the player will get at most 26 hits in 100 times at bat.

Solution

Here, $p = 0.32$, $q = 0.68$, and $n = 100$.

Step 1 Since $np = (100)(0.320) = 32$ and $nq = (100)(0.680) = 68$, the normal distribution can be used to approximate the binomial distribution.

Step 2 $\mu = np = (100)(0.320) = 32$

$$\sigma = \sqrt{npq} = \sqrt{(100)(0.32)(0.68)} = \sqrt{21.76} = 4.66$$

Step 3 $P(X \leq 26)$.

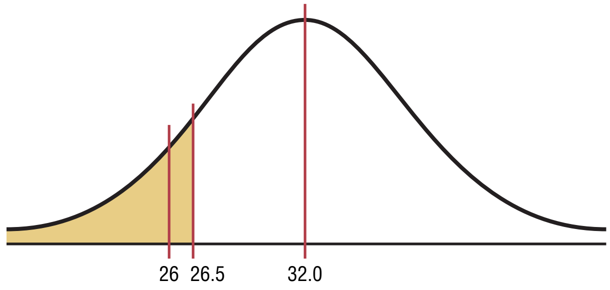
Step 4 See approximation number 4 in Table 6–2: $P(X < 26 + 0.5) = P(X < 26.5)$. The desired area is shown in Figure 6–50.

Step 5 The z value is

$$z = \frac{26.5 - 32}{4.66} = -1.18$$

Figure 6–50

Area Under a Normal Curve for Example 6–26



Step 6 The area between the mean and 26.5 is 0.3810. Since the area in the left tail is desired, 0.3810 must be subtracted from 0.5000. So the probability is $0.5000 - 0.3810 = 0.1190$, or 11.9%.

The closeness of the normal approximation is shown in Example 6–27.

Example 6–27

When $n = 10$ and $p = 0.5$, use the binomial distribution table (Table B in Appendix C) to find the probability that $X = 6$. Then use the normal approximation to find the probability that $X = 6$.

Solution

From Table B, for $n = 10$, $p = 0.5$, and $X = 6$, the probability is 0.205.

For a normal approximation,

$$\mu = np = (10)(0.5) = 5$$

$$\sigma = \sqrt{npq} = \sqrt{(10)(0.5)(0.5)} = 1.58$$

Now, $X = 6$ is represented by the boundaries 5.5 and 6.5. So the z values are

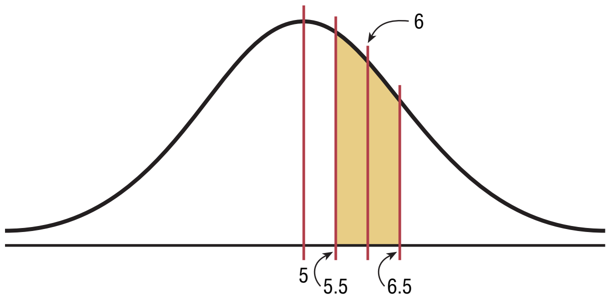
$$z_1 = \frac{6.5 - 5}{1.58} = 0.95 \qquad z_2 = \frac{5.5 - 5}{1.58} = 0.32$$

The corresponding area for 0.95 is 0.3289, and the corresponding area for 0.32 is 0.1255.

The solution is $0.3289 - 0.1255 = 0.2034$, which is very close to the binomial table value of 0.205. The desired area is shown in Figure 6–51.

Figure 6–51

Area Under a Normal Curve for Example 6–27



The normal approximation also can be used to approximate other distributions, such as the Poisson distribution (see Table C in Appendix C).

Applying the Concepts 6–6

How Safe Are You?

Assume one of your favorite activities is mountain climbing. When you go mountain climbing, you have several safety devices to keep you from falling. You notice that attached to one of your safety hooks is a reliability rating of 97%. You estimate that throughout the next year you will be using this device about 100 times. Answer the following questions.

1. Does a reliability rating of 97% mean that there is a 97% chance that the device will not fail any of the 100 times?
2. What is the probability of at least one failure?
3. What is the complement of this event?
4. Can this be considered a binomial experiment?
5. Can you use the binomial probability formula? Why or why not?
6. Find the probability of at least two failures.
7. Can you use a normal distribution to accurately approximate the binomial distribution? Explain why or why not.
8. Is correction for continuity needed?
9. How much safer would it be to use a second safety hook independently of the first?

See page 345 for the answers.

Exercises 6–6

1. Explain why a normal distribution can be used as an approximation to a binomial distribution. What conditions must be met to use the normal distribution to approximate the binomial distribution? Why is a correction for continuity necessary?

2 (ans) Use the normal approximation to the binomial to find the probabilities for the specific value(s) of X.

a. $n = 30, p = 0.5, X = 18$
b. $n = 50, p = 0.8, X = 44$
c. $n = 100, p = 0.1, X = 12$
d. $n = 10, p = 0.5, X \geq 7$
e. $n = 20, p = 0.7, X \leq 12$
f. $n = 50, p = 0.6, X \leq 40$

3. Check each binomial distribution to see whether it can be approximated by a normal distribution (i.e., are $np \geq 5$ and $nq \geq 5$?).

a. $n = 20, p = 0.5$
b. $n = 10, p = 0.6$
c. $n = 40, p = 0.9$
d. $n = 50, p = 0.2$
e. $n = 30, p = 0.8$
f. $n = 20, p = 0.85$

4. Of all 3- to 5-year-old children, 56% are enrolled in school. If a sample of 500 such children is randomly selected, find the probability that at least 250 will be enrolled in school.

Source: *Statistical Abstract of the United States.*

5. Two out of five adult smokers acquired the habit by age 14. If 400 smokers are randomly selected, find the probability that 170 or more acquired the habit by age 14.

Source: *Harper's Index.*

6. A theater owner has found that 5% of patrons do not show up for the performance that they purchased tickets for. If the theater has 100 seats, find the probability that six or more patrons will not show up for the sold-out performance.

7. The percentage of Americans 25 years or older who have at least some college education is 50.9%. In a random sample of 300 Americans 25 years old and older, what is the probability that more than 175 have at least some college education?

Source: *N.Y. Times Almanac.*

8. According to recent surveys, 53% of households have personal computers. If a random sample of 175 households is selected, what is the probability that more than 75 but fewer than 110 have a personal computer?

Source: *N.Y. Times Almanac.*

9. The percentage of female Americans 25 years old and older who have completed 4 years of college or more is 23.6%. In a random sample of 180, American women who are at least 25, what is the probability that more than 50 have completed 4 years of college or more?

Source: *N.Y. Times Almanac.*

10. Women make up 24% of the science and engineering workforce. In a random sample of 400 science and engineering employees, what is the probability that more than 120 are women?

Source: *Science and Engineering Indicators,* www.nsf.gov.

11. Women comprise 83.3% of all elementary school teachers. In a random sample of 300 elementary school teachers, what is the probability that more than 50 are men?

Source: *N.Y. Times Almanac.*

12. Seventy-seven percent of U.S. homes have a telephone answering device. In a random sample of 290 homes, what is the probability that more than 50 do not have an answering device?

Source: *N.Y. Times Almanac.*

13. The mayor of a small town estimates that 35% of the residents in his town favor the construction of a municipal parking lot. If there are 350 people at a town meeting, find the probability that at least 100 favor construction of the parking lot. Based on your answer, is it likely that 100 or more people would favor the parking lot?

Extending the Concepts

14. Recall that for use of a normal distribution as an approximation to the binomial distribution, the conditions $np \geq 5$ and $nq \geq 5$ must be met. For each given probability, compute the minimum sample size needed for use of the normal approximation.

a. $p = 0.1$
b. $p = 0.3$
c. $p = 0.5$
d. $p = 0.8$
e. $p = 0.9$

6–7 Summary

A normal distribution can be used to describe a variety of variables, such as heights, weights, and temperatures. A normal distribution is bell-shaped, unimodal, symmetric, and continuous; its mean, median, and mode are equal. Since each variable has its own distribution with mean μ and standard deviation σ, mathematicians use the standard normal distribution, which has a mean of 0 and a standard deviation of 1. Other approximately normally distributed variables can be transformed to the standard normal distribution with the formula $z = (X - \mu)/\sigma$.

A normal distribution can also be used to describe a sampling distribution of sample means. These samples must be of the same size and randomly selected with replacement from the population. The means of the samples will differ somewhat from the population mean, since samples are generally not perfect representations of the population from which they came. The mean of the sample means will be equal to the population mean; and the standard deviation of the sample means will be equal to the population standard deviation, divided by the square root of the sample size. The central limit theorem states that as the size of the samples increases, the distribution of sample means will be approximately normal.

A normal distribution can be used to approximate other distributions, such as a binomial distribution. For a normal distribution to be used as an approximation, the conditions $np \geq 5$ and $nq \geq 5$ must be met. Also, a correction for continuity may be used for more accurate results.

Important Terms

central limit theorem 324

correction for continuity 333

negatively or left-skewed distribution 287

normal distribution 289

positively or right-skewed distribution 287

sampling distribution of sample means 322

sampling error 322

standard error of the mean 324

standard normal distribution 290

symmetric distribution 287

z value 290

Important Formulas

Formula for the z value (or standard score):

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

Formula for finding a specific data value:

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

Formula for the mean of the sample means:

$$\mu_{\bar{X}} = \mu$$

Formula for the standard error of the mean:

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

Formula for the z value for the central limit theorem:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Formulas for the mean and standard deviation for the binomial distribution:

$$\mu = n \cdot p \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

Review Exercises

1. Find the area under the standard normal distribution curve for each.

a. Between $z = 0$ and $z = 1.95$
b. Between $z = 0$ and $z = 0.37$
c. Between $z = 1.32$ and $z = 1.82$
d. Between $z = -1.05$ and $z = 2.05$
e. Between $z = -0.03$ and $z = 0.53$
f. Between $z = +1.10$ and $z = -1.80$
g. To the right of $z = 1.99$
h. To the right of $z = -1.36$
i. To the left of $z = -2.09$
j. To the left of $z = 1.68$

2. Using the standard normal distribution, find each probability.

a. $P(0 < z < 2.07)$
b. $P(-1.83 < z < 0)$
c. $P(-1.59 < z < +2.01)$
d. $P(1.33 < z < 1.88)$
e. $P(-2.56 < z < 0.37)$
f. $P(z > 1.66)$
g. $P(z < -2.03)$
h. $P(z > -1.19)$
i. $P(z < 1.93)$
j. $P(z > -1.77)$

3. If the mean salary of auto mechanics in the United States is \$27,635 and the standard deviation is \$2550, find these probabilities for a randomly selected auto mechanic. Assume the variable is normally distributed.

a. The mechanic earns more than \$27,635.
b. The mechanic earns less than \$25,000.
c. If you were offered a job at \$25,000, how would you compare your salary with the salaries of the population of auto mechanics?

4. The average salary for graduates entering the actuarial field is \$40,000. If the salaries are normally distributed with a standard deviation of \$5000, find the probability that

a. An individual graduate will have a salary over \$45,000.
b. A group of nine graduates will have a group average over \$45,000.

Source: www.BeAnActuary.org.

5. The speed limit on Interstate 75 around Findlay, Ohio, is 65 mph. On a clear day with no construction, the mean speed of automobiles was measured at 63 mph with a standard deviation of 8 mph. If the speeds are normally distributed, what percentage of the automobiles are exceeding the speed limit? If the Highway Patrol decides to ticket only motorists exceeding 72 mph, what percentage of the motorists might they arrest?

6. The national average for a new car loan was 8.28%. If the rate is normally distributed with a standard deviation of 3.5%, find these probabilities.

a. One can receive a rate less than 9%.
b. One can receive a rate less than 8%.

Source: www.bankrate.com/New York Times.

7. For the first 7 months of the year, the average precipitation in Toledo, Ohio, is 19.32 inches. If the average precipitation is normally distributed with a standard deviation of 2.44 inches, find these probabilities.

a. A randomly selected year will have precipitation greater than 18 inches for the first 7 months.
b. Five randomly selected years will have an average precipitation greater than 18 inches for the first 7 months.

Source: Toledo Blade.

8. The average weight of an airline passenger's suitcase is 45 pounds. The standard deviation is 2 pounds. If 15% of the suitcases are overweight, find the maximum weight allowed by the airline. Assume the variable is normally distributed.

9. An educational study to be conducted requires a test score in the middle 40% range. If $\mu = 100$ and $\sigma = 15$, find the highest and lowest acceptable test scores that would enable a candidate to participate in the study. Assume the variable is normally distributed.

10. The average cost of XYZ brand running shoes is \$83 per pair, with a standard deviation of \$8. If nine pairs of running shoes are selected, find the probability that the mean cost of a pair of shoes will be less than \$80. Assume the variable is normally distributed.

11. A recent study of the life span of portable compact disc players found the average to be 3.7 years with a standard deviation of 0.6 year. If a random sample of 32 people who own CD players is selected, find the probability that the mean lifetime of the sample will be less than 3.4 years. If the mean is less than 3.4 years, would you consider that 3.7 years might be incorrect?

12. The probability of winning on a slot machine is 5%. If a person plays the machine 500 times, find the probability of winning 30 times. Use the normal approximation to the binomial distribution.

13. Of the total population of older Americans, 18% live in Florida. For a randomly selected sample of 200 older Americans, find the probability that more than 40 live in Florida.

Source: Elizabeth Vierck, *Fact Book on Aging.*

14. In a large university, 30% of the incoming freshmen elect to enroll in a personal finance course offered by the university. Find the probability that of 800 randomly selected incoming freshmen, at least 260 have elected to enroll in the course.

15. Of the total population of the United States, 20% live in the northeast. If 200 residents of the United States are selected at random, find the probability that at least 50 live in the northeast.

Source: Statistical Abstract of the United States.

16. The heights (in feet above sea level) of a random sample of the world's active volcanoes are shown here. Check for normality.

13,435	5,135	11,339	12,224	7,470
9,482	12,381	7,674	5,223	5,631
3,566	7,113	5,850	5,679	15,584
5,587	8,077	9,550	8,064	2,686
5,250	6,351	4,594	2,621	9,348
6,013	2,398	5,658	2,145	3,038

Source: N.Y. Times Almanac.

17. A random sample of enrollments in Pennsylvania's private four-year colleges is listed here. Check for normality.

1350	1886	1743	1290	1767
2067	1118	3980	1773	4605
1445	3883	1486	980	1217
3587				

Source: N.Y. Times Almanac

Statistics Today

What Is Normal?—Revisited

Many of the variables measured in medical tests—blood pressure, triglyceride level, etc.—are approximately normally distributed for the majority of the population in the United States. Thus, researchers can find the mean and standard deviation of these variables. Then, using these two measures along with the z values, they can find normal intervals for healthy individuals. For example, 95% of the systolic blood pressures of healthy individuals fall within 2 standard deviations of the mean. If an individual's pressure is outside the determined normal range (either above or below), the physician will look for a possible cause and prescribe treatment if necessary.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. The total area under a normal distribution is infinite.
2. The standard normal distribution is a continuous distribution.
3. All variables that are approximately normally distributed can be transformed to standard normal variables.
4. The z value corresponding to a number below the mean is always negative.
5. The area under the standard normal distribution to the left of $z = 0$ is negative.
6. The central limit theorem applies to means of samples selected from different populations.

Select the best answer.

7. The mean of the standard normal distribution is

   a. 0 *c.* 100
   b. 1 *d.* Variable

8. Approximately what percentage of normally distributed data values will fall within 1 standard deviation above or below the mean?

   a. 68%
   b. 95%
   c. 99.7%
   d. Variable

9. Which is not a property of the standard normal distribution?

   a. It's symmetric about the mean.
   b. It's uniform.
   c. It's bell-shaped.
   d. It's unimodal.

10. When a distribution is positively skewed, the relationship of the mean, median, and mode from left to right will be

    a. Mean, median, mode
    b. Mode, median, mean
    c. Median, mode, mean
    d. Mean, mode, median

11. The standard deviation of all possible sample means equals

    a. The population standard deviation.
    b. The population standard deviation divided by the population mean.
    c. The population standard deviation divided by the square root of the sample size.
    d. The square root of the population standard deviation.

Complete the following statements with the best answer.

12. When one is using the standard normal distribution, $P(z < 0) =$ ________.
13. The difference between a sample mean and a population mean is due to ________.
14. The mean of the sample means equals ________.
15. The standard deviation of all possible sample means is called ________.
16. The normal distribution can be used to approximate the binomial distribution when $n \cdot p$ and $n \cdot q$ are both greater than or equal to ________.
17. The correction factor for the central limit theorem should be used when the sample size is greater than ________ the size of the population.
18. Find the area under the standard normal distribution for each.

    a. Between 0 and 1.50
    b. Between 0 and −1.25
    c. Between 1.56 and 1.96
    d. Between −1.20 and −2.25
    e. Between −0.06 and 0.73
    f. Between 1.10 and −1.80
    g. To the right of $z = 1.75$
    h. To the right of $z = -1.28$
    i. To the left of $z = -2.12$
    j. To the left of $z = 1.36$

19. Using the standard normal distribution, find each probability.

    a. $P(0 < z < 2.16)$
    b. $P(-1.87 < z < 0)$
    c. $P(-1.63 < z < 2.17)$
    d. $P(1.72 < z < 1.98)$
    e. $P(-2.17 < z < 0.71)$
    f. $P(z > 1.77)$
    g. $P(z < -2.37)$
    h. $P(z > -1.73)$
    i. $P(z < 2.03)$
    j. $P(z > -1.02)$

20. The average amount of rain per year in Greenville is 49 inches. The standard deviation is 8 inches. Find the probability that next year Greenville will receive the following amount of rainfall. Assume the variable is normally distributed.

    a. At most 55 inches of rain
    b. At least 62 inches of rain
    c. Between 46 and 54 inches of rain
    d. How many inches of rain would you consider to be an extremely wet year?

21. The average height of a certain age group of people is 53 inches. The standard deviation is 4 inches. If the

variable is normally distributed, find the probability that a selected individual's height will be

a. Greater than 59 inches.
b. Less than 45 inches.
c. Between 50 and 55 inches.
d. Between 58 and 62 inches.

22. The average number of gallons of lemonade consumed by the football team during a game is 20, with a standard deviation of 3 gallons. Assume the variable is normally distributed. When a game is played, find the probability of using

a. Between 20 and 25 gallons.
b. Less than 19 gallons.
c. More than 21 gallons.
d. Between 26 and 28 gallons.

23. The average number of years a person takes to complete a graduate degree program is 3. The standard deviation is 4 months. Assume the variable is normally distributed. If an individual enrolls in the program, find the probability that it will take

a. More than 4 years to complete the program.
b. Less than 3 years to complete the program.
c. Between 3.8 and 4.5 years to complete the program.
d. Between 2.5 and 3.1 years to complete the program.

24. On the daily run of an express bus, the average number of passengers is 48. The standard deviation is 3. Assume the variable is normally distributed. Find the probability that the bus will have

a. Between 36 and 40 passengers.
b. Fewer than 42 passengers.
c. More than 48 passengers.
d. Between 43 and 47 passengers.

25. The average thickness of books on a library shelf is 8.3 centimeters. The standard deviation is 0.6 centimeter. If 20% of the books are oversized, find the minimum thickness of the oversized books on the library shelf. Assume the variable is normally distributed.

26. Membership in an elite organization requires a test score in the upper 30% range. If $\mu = 115$ and $\sigma = 12$, find the lowest acceptable score that would enable a candidate to apply for membership. Assume the variable is normally distributed.

27. The average repair cost of a microwave oven is \$55, with a standard deviation of \$8. The costs are normally distributed. If 12 ovens are repaired, find the probability that the mean of the repair bills will be greater than \$60.

28. The average electric bill in a residential area is \$72 for the month of April. The standard deviation is \$6. If the amounts of the electric bills are normally distributed, find the probability that the mean of the bill for 15 residents will be less than \$75.

29. According to a recent survey, 38% of Americans get 6 hours or less of sleep each night. If 25 people are selected, find the probability that 14 or more people will get 6 hours or less of sleep each night. Does this number seem likely?

Source: Amazing Almanac.

30. If 10% of the people in a certain factory are members of a union, find the probability that, in a sample of 2000, fewer than 180 people are union members.

31. The percentage of U.S. households that have online connections is 44.9%. In a random sample of 420 households, what is the probability that fewer than 200 have online connections?

Source: N.Y. Times Almanac.

32. Fifty-three percent of U.S. households have a personal computer. In a random sample of 250 households, what is the probability that fewer than 120 have a PC?

Source: N.Y. Times Almanac.

33. The number of calories contained in a selection of fast-food sandwiches is shown here. Check for normality.

390	405	580	300	320
540	225	720	470	560
535	660	530	290	440
390	675	530	1010	450
320	460	290	340	610
430	530			

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

34. The average GMAT scores for the top-30 ranked graduate schools of business are listed here. Check for normality.

718	703	703	703	700	690	695	705	690	688
676	681	689	686	691	669	674	652	680	670
651	651	637	662	641	645	645	642	660	636

Source: *U.S. News & World Report Best Graduate Schools.*

Critical Thinking Challenges

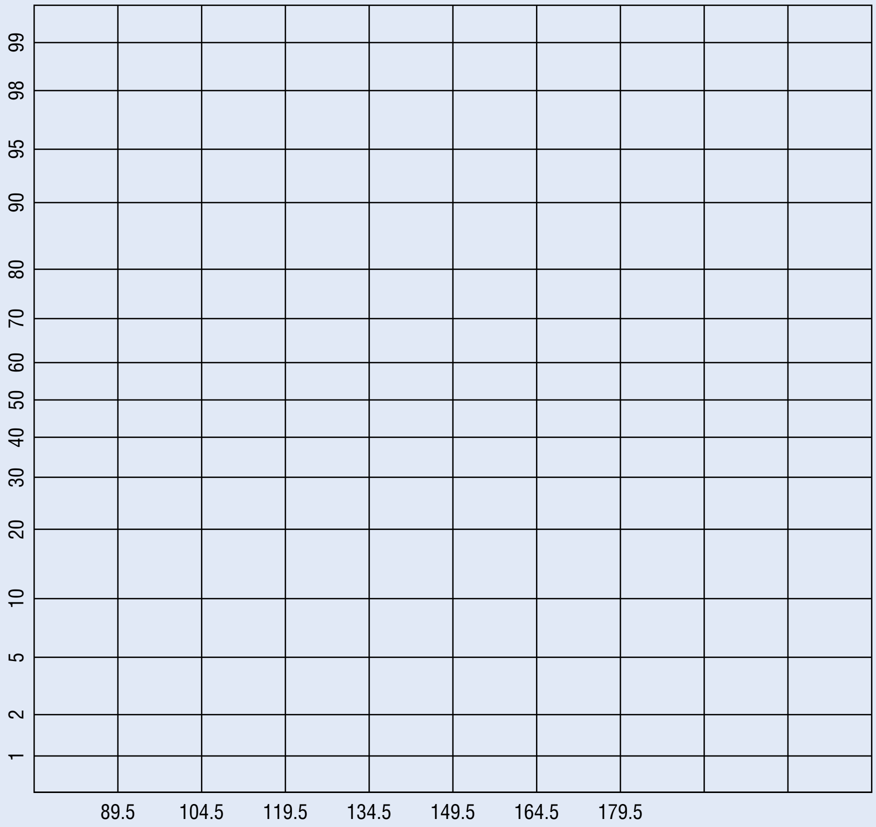
Sometimes a researcher must decide whether a variable is normally distributed. There are several ways to do this. One simple but very subjective method uses special graph paper, which is called *normal probability paper.* For the distribution of systolic blood pressure readings given in Chapter 3 of the textbook, the following method can be used:

1. Make a table, as shown.

Boundaries	Frequency	Cumulative frequency	Cumulative percent frequency
89.5–104.5	24		
104.5–119.5	62		
119.5–134.5	72		
134.5–149.5	26		
149.5–164.5	12		
164.5–179.5	4		
	200		

2. Find the cumulative frequencies for each class, and place the results in the third column.
3. Find the cumulative percents for each class by dividing each cumulative frequency by 200 (the total frequencies) and multiplying by 100%. (For the first class, it would be $24/200 \times 100\% = 12\%$.) Place these values in the last column.
4. Using the normal probability paper shown in Table 6–3, label the x axis with the class boundaries as shown and plot the percents.

Table 6–3 Normal Probability Paper



5. If the points fall approximately in a straight line, it can be concluded that the distribution is normal. Do you feel that this distribution is approximately normal? Explain your answer.
6. To find an approximation of the mean or median, draw a horizontal line from the 50% point on the y axis over to the curve and then a vertical line down to the x axis. Compare this approximation of the mean with the computed mean.
7. To find an approximation of the standard deviation, locate the values on the x axis that correspond to the 16 and 84% values on the y axis. Subtract these two values and divide the result by 2. Compare this approximate standard deviation to the computed standard deviation.
8. Explain why the method used in step 7 works.

Data Projects

1. Select a variable (interval or ratio) and collect 30 data values. Some suggestions might include heights of the students in your class, grade point averages of students, running times of feature length movies, etc.
   a. Construct a frequency distribution for the variable.
   b. Use the procedure described in the Critical Thinking Challenge on page 342 to graph the distribution on normal probability paper.
   c. Can you conclude that the data are approximately normally distributed? Explain your answer.
2. Repeat Exercise 1, using all the values from Data Set II in Appendix D.
3. Repeat Exercise 1, using a random sample of 30 values from Data Set III in Appendix D.

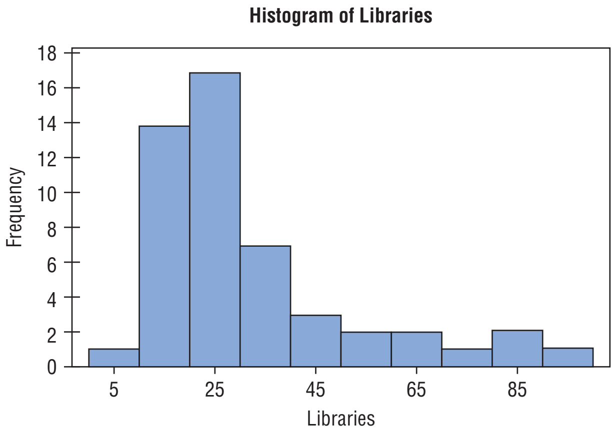
Answers to Applying the Concepts

Section 6–3 Assessing Normality

1. Answers will vary. One possible frequency distribution is the following:

Branches	Frequency
0–9	1
10–19	14
20–29	17
30–39	7
40–49	3
50–59	2
60–69	2
70–79	1
80–89	2
90–99	1

2. Answers will vary according to the frequency distribution in question 1. This histogram matches the frequency distribution in question 1.



3. The histogram is unimodal and skewed to the right (positively skewed).
4. The distribution does not appear to be normal.
5. The mean number of branches is $\bar{x} = 31.4$, and the standard deviation is $s = 20.6$.
6. Of the data values, 80% fall within 1 standard deviation of the mean (between 10.8 and 52).
7. Of the data values, 92% fall within 2 standard deviations of the mean (between 0 and 72.6).
8. Of the data values, 98% fall within 3 standard deviations of the mean (between 0 and 93.2).
9. My values in questions 6–8 differ from the 68, 95, and 100% that we would see in a normal distribution.
10. These values support the conclusion that the distribution of the variable is not normal.

Section 6–4 Smart People

1. $z = \frac{130 - 100}{15} = 2$. The area to the right of 2 in the standard normal table is about 0.0228, so I would expect about 10,000(0.0228) = 228 people in Visiala to qualify for Mensa.
2. It does seem reasonable to continue my quest to start a Mensa chapter in Visiala.
3. Answers will vary. One possible answer would be to randomly call telephone numbers (both home and cell phones) in Visiala, ask to speak to an adult, and ask whether the person would be interested in joining Mensa.
4. To have an Ultra-Mensa club, I would need to find the people in Visiala who have IQs that are at least 2.326 standard deviations above average. This means that I would need to recruit those with IQs that are at least 135:

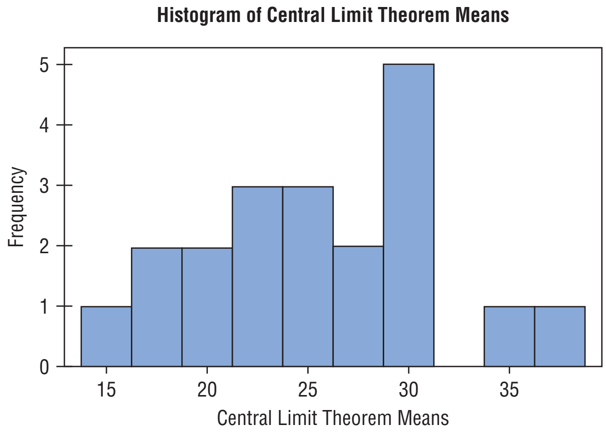
$$2.326 = \frac{x - 100}{15} \Rightarrow x = 100 + 2.326(15) = 134.89$$

Section 6–5 Central Limit Theorem

1. It is very unlikely that we would ever get the same results for any of our random samples. While it is a remote possibility, it is highly unlikely.
2. A good estimate for the population mean would be to find the average of the students' sample means.

Similarly, a good estimate for the population standard deviation would be to find the average of the students' sample standard deviations.

3. The distribution appears to be somewhat left (negatively) skewed.



4. The mean of the students' means is 25.4, and the standard deviation is 5.8.

5. The distribution of the means is not a sampling distribution, since it represents just 20 of all possible samples of size 30 from the population.

6. The sampling error for student 3 is $18 - 25.4 = -7.4$; the sampling error for student 7 is $26 - 25.4 = +0.6$; the sampling error for student 14 is $29 - 25.4 = +3.6$.

7. The standard deviation for the sample of the 20 means is greater than the standard deviations for each of the individual students. So it is not equal to the standard deviation divided by the square root of the sample size.

Section 6–6 How Safe Are You?

1. A reliability rating of 97% means that, on average, the device will not fail 97% of the time. We do not know how many times it will fail for any particular set of 100 climbs.

2. The probability of at least 1 failure in 100 climbs is $1 - (0.97)^{100} = 1 - 0.0476 = 0.9524$ (about 95%).

3. The complement of the event in question 2 is the event of "no failures in 100 climbs."

4. This can be considered a binomial experiment. We have two outcomes: success and failure. The probability of the equipment working (success) remains constant at 97%. We have 100 independent climbs. And we are counting the number of times the equipment works in these 100 climbs.

5. We could use the binomial probability formula, but it would be very messy computationally.

6. The probability of at least two failures *cannot* be estimated with the normal distribution (see below). So the probability is $1 - [(0.97)^{100} + 100(0.97)^{99}(0.03)] = 1 - 0.1946 = 0.8054$ (about 80.5%).

7. We *should not* use the normal approximation to the binomial since $nq < 10$.

8. If we had used the normal approximation, we would have needed a correction for continuity, since we would have been approximating a discrete distribution with a continuous distribution.

9. Since a second safety hook will be successful or fail independently of the first safety hook, the probability of failure drops from 3% to $(0.03)(0.03) = 0.0009$, or 0.09%.