7

Normal Curves and Sampling Distributions

One cannot escape the feeling that these mathematical formulas have an independent existence and an intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than was originally put into them.

—Heinrich Hertz

How can it be that mathematics, a product of human thought independent of experience, is so admirably adapted to the objects of reality?

—Albert Einstein

Heinrich Rudolf Hertz (1857–1894)

This German physicist was largely responsible for doing pioneering work in electromagnetic theory.

Heinrich Hertz was a pioneer in the study of radio waves. His work and the later work of Maxwell and Marconi led the way to modern radio, television, and radar. Albert Einstein is world renowned for his great discoveries in relativity and quantum mechanics. Everyone who has worked in both mathematics and real-world applications cannot help but marvel at how the "pure thought" of the mathematical sciences can predict and explain events in other realms. In this chapter, we will study the most important type of probability distribution in all of mathematical statistics: the normal distribution. Why is the normal distribution so important? Two of the reasons are that it applies to a wide variety of situations and that other distributions tend to become normal under certain conditions.

PREVIEW QUESTIONS

◇ What are some characteristics of a normal distribution? What does the empirical rule tell you about data spread around the mean? (SECTION 7.1)

◇ Can you compare apples and oranges, or maybe elephants and butterflies? In most cases, the answer is no . . . unless you first *standardize* your measurements. What are a *standard normal distribution* and a *standard z score*? (SECTION 7.2)

◇ How do you convert *any* normal distribution to a *standard normal* distribution? How do you find probabilities of "standardized events"? (SECTION 7.3)

For on-line student resources, visit **math.college.hmco.com/students** and follow the Statistics links to the Brase/Brase, *Understanding Basic Statistics*, 4th edition web site.

◇ As humans, our experiences are finite and limited. Consequently, most of the important decisions in our lives are based on sample (incomplete) information. What is a probability sampling distribution? How will sampling distributions help us make good decisions based on incomplete information? (SECTION 7.4)

◇ There is an old saying: All roads lead to Rome. In statistics, we could recast this saying: All probability distributions average out to be normal distributions (as the sample size increases). How can we take advantage of this in our study of sampling distributions? (SECTION 7.5)

◇ The binomial and normal distributions are two of the most important probability distributions in statistics. Under certain limiting conditions, the binomial can be thought to evolve (or envelope) into the normal distribution. How can you apply this in the real world? (SECTION 7.6)

FOCUS PROBLEM

Impulse Buying

The Food Marketing Institute, Progressive Grocer, New Products News, and Point of Purchaser Advertising Institute are organizations that analyze supermarket sales. One of the interesting discoveries was that the average amount of impulse buying in a grocery store was very time-dependent.

As reported in the *Denver Post,* "when you dilly dally in a store for 10 unplanned minutes, you can kiss nearly \$20 goodbye." For this reason, it is in the best interest of the supermarket to keep you in the store longer. In the *Post* article, it was pointed out that long checkout lines (near end-aisle displays), "samplefest" events of tasting free samples, video kiosks, magazine and book sections, and so on help keep customers in the store longer. On average, a single customer who strays from his or her grocery list can plan on impulse spending of \$20 for every 10 minutes spent wandering about in the supermarket.

Let x represent the dollar amount spent on supermarket impulse buying in a 10-minute (unplanned) shopping interval. Based on the *Post* article, the mean of the x distribution is about \$20 and the (estimated) standard deviation is about \$7.

(a) Let us assume that x has a distribution that is approximately normal. What is the probability that x is between \$18 and \$22?

(b) Consider a random sample of $n = 100$ customers, each of whom has 10 minutes of unplanned shopping time in a supermarket. From the central limit theorem, what can you say about the probability distribution of $\bar{x}$, the *average* amount spent by these customers due to impulse buying? Is the $\bar{x}$ distribution approximately normal? What are the mean and standard deviation of the $\bar{x}$ distribution? Is it necessary to make any assumption about the x distribution? Explain.

(c) What is the probability that $\bar{x}$ is between \$18 and \$22?

(d) In part (c), we used $\bar{x}$, the *average* amount spent, computed for 100 customers. In part (a), we used x, the amount spent by only *one* individual customer. The answers for parts (c) and (a) are very different. Why would this happen? In this example, $\bar{x}$ is a much more predictable or reliable statistic than x. Consider that almost all marketing strategies and sales pitches are designed for the *average* customer and *not* the *individual* customer. How does the central limit theorem tell us that the average customer is much more predictable than the individual customer? (See Problem 12 of Section 7.5.)

7.1 Graphs of Normal Probability Distributions

FOCUS POINTS

- ✓ Graph a normal curve and summarize its important properties.
- ✓ Apply the empirical rule to solve real-world problems.

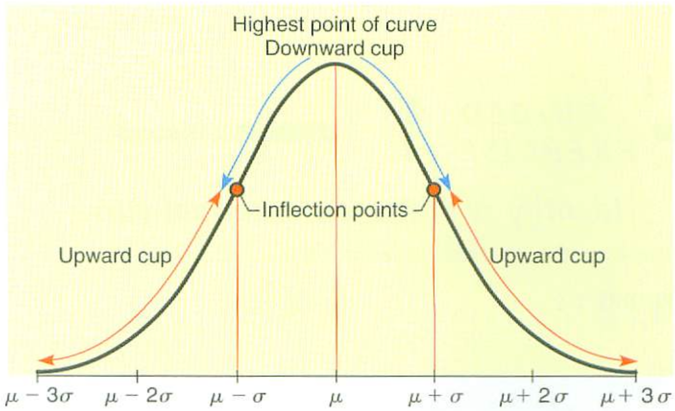
One of the most important examples of a continuous probability distribution is the *normal distribution.* This distribution was studied by the French mathematician Abraham de Moivre (1667–1754) and later by the German mathematician Carl Friedrich Gauss (1777–1855), whose work is so important that the normal distribution is sometimes called *Gaussian.* The work of these mathematicians provided a foundation on which much of the theory of statistical inference is based.

Applications of a normal probability distribution are so numerous that some mathematicians refer to it as "a veritable Boy Scout knife of statistics." However, before we can apply it, we must examine some of the properties of a normal distribution.

A rather complicated formula, presented later in this section, defines a normal distribution in terms of μ and σ, the mean and standard deviation of the population

FIGURE 7-1

A Normal Curve



distribution. It is only through this formula that we can verify if a distribution is normal. However, we can look at the graph of a normal distribution and get a good pictorial idea of some of the essential features of any normal distribution.

Normal curve

The graph of a normal distribution is called a *normal curve*. It possesses a shape very much like the cross section of a pile of dry sand. Because of its shape, blacksmiths would sometimes use a pile of dry sand in the construction of a mold for a bell. Thus the normal curve is also called a *bell-shaped curve* (see Figure 7-1).

We see that a general normal curve is smooth and symmetrical about the vertical line extending upward from the mean μ. Notice that the highest point of the curve occurs over μ. If the distribution were graphed on a piece of sheet metal, cut out, and placed on a knife edge, the balance point would be at μ. We also see that the curve tends to level out and approach the horizontal (x axis) like a glider making a landing. However, in mathematical theory, such a glider would never quite finish its landing because a normal curve never touches the horizontal axis.

The parameter σ controls the spread of the curve. The curve is quite close to the horizontal axis at $\mu + 3\sigma$ and $\mu - 3\sigma$. Thus, if the standard deviation σ is large, the curve will be more spread out; if it is small, the curve will be more peaked. Figure 7-1 shows the normal curve cupped downward for an interval on either side of the mean μ. Then it begins to cup upward as we go to the lower part of the bell. The exact places where the *transition* between the upward and downward cupping occurs are above the points $\mu + \sigma$ and $\mu - \sigma$. In the terminology of calculus, transition points such as these are called *inflection points*.

Important properties of a normal curve

1. The curve is bell-shaped with the highest point over the mean μ.
2. The curve is symmetrical about a vertical line through μ.
3. The curve approaches the horizontal axis but never touches or crosses it.
4. The inflection (transition) points between cupping upward and downward occur above $\mu + \sigma$ and $\mu - \sigma$.

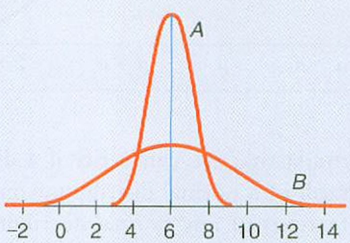
The parameters that control the shape of a normal curve are the mean μ and the standard deviation σ. When both μ and σ are specified, a specific normal curve is determined. In brief, μ locates the balance point, and σ determines the extent of the spread.

GUIDED EXERCISE 1

Identify μ and σ on a normal curve

Look at the normal curves in Figure 7-2.

FIGURE 7-2



(a) Do these distributions have the same mean? If so, what is it?

The means are the same, since both graphs have the high point over 6. $\mu = 6$.

(b) One of the curves corresponds to a normal distribution with $\sigma = 3$ and the other to one with $\sigma = 1$. Which curve has which σ?

Curve A has $\sigma = 1$, and curve B has $\sigma = 3$. (Since curve B is more spread out, it has the larger σ value.)

◇ **COMMENT** The normal distribution curve is always above the horizontal axis. The area beneath the curve and above the axis is exactly one. As such, the normal distribution curve is an example of a *density curve*. The formula used to generate the shape of the normal distribution curve is called the *normal density function*. If x is a normal random variable with mean μ and standard deviation σ, the formula for the normal density function is

$$f(x) = \frac{e^{(-1/2)((x-\mu)/\sigma)^2}}{\sigma\sqrt{2\pi}}$$

In this text, we will not use this formula explicitly. However, we will use tables of areas based on the normal density function. ◇

The total area under any normal curve studied in this book will *always* be 1. The graph of the normal distribution is important because the portion of the *area* under the curve above a given interval represents the *probability* that a measurement will lie in that interval.

In Section 3.2, we studied Chebyshev's theorem. This theorem gives us information about the *smallest* proportion of data that lies within 2, 3, or k standard deviations of the mean. This result applies to *any* distribution. However, for normal distributions, we can get a much more precise result, which is given by the *empirical rule*.

Empirical rule

Empirical rule

For a distribution that is symmetrical and bell-shaped (in particular, for a normal distribution):

Approximately 68% of the data values will lie within one standard deviation on each side of the mean.

Approximately 95% of the data values will lie within two standard deviations on each side of the mean.

Approximately 99.7% (or almost all) of the data values will lie within three standard deviations on each side of the mean.

The preceding statement is called the *empirical rule* because, for symmetrical, bell-shaped distributions, the given percentages are observed in practice. Furthermore, for the normal distribution, the empirical rule is a direct consequence of the very nature of the distribution (see Figure 7-3). Notice that the empirical rule is a stronger statement than Chebyshev's theorem in that it gives *definite percentages,* not just lower limits. Of course, the empirical rule applies only to normal or symmetrical, bell-shaped distributions, whereas Chebyshev's theorem applies to all distributions.

EXAMPLE 1
Empirical rule

The playing life of a Sunshine radio is normally distributed with mean $\mu = 600$ hours and standard deviation $\sigma = 100$ hours. What is the probability that a radio selected at random will last from 600 to 700 hours?

SOLUTION: The probability that the playing time will be between 600 and 700 hours is equal to the percentage of the total area under the curve that is shaded in Figure 7-4. Since $\mu = 600$ and $\mu + \sigma = 600 + 100 = 700$, we see that the shaded area is simply the area between μ and $\mu + \sigma$. The area from μ to $\mu + \sigma$ is 34% of the total area. This tells us that the probability a Sunshine radio will last between 600 and 700 playing hours is about 0.34. ◊

FIGURE 7-3

Area Under a Normal Curve

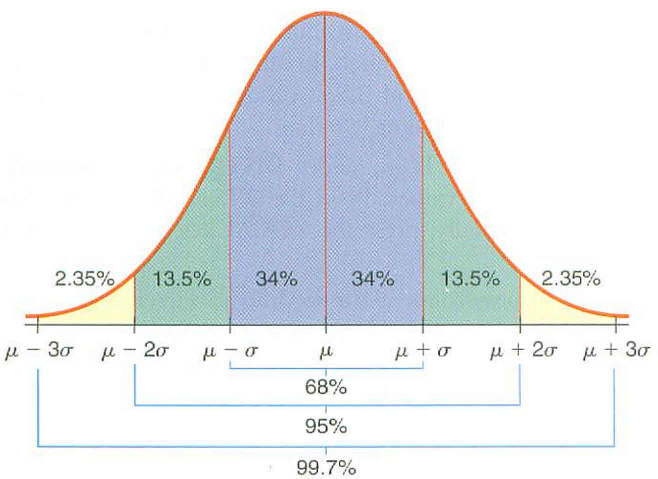
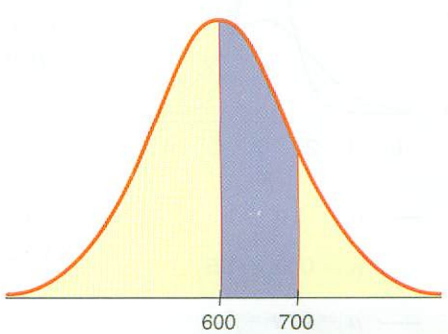


FIGURE 7-4

Distribution of Playing Times



GUIDED EXERCISE 2

Empirical rule

The yearly wheat yield per acre on a particular farm is normally distributed with mean $\mu = 35$ bushels and standard deviation $\sigma = 8$ bushels.

(a) Shade the area under the curve in Figure 7-5 that represents the probability that an acre will yield between 19 and 35 bushels.

→ See Figure 7-6.

(b) Is the area the same as the area between $\mu - 2\sigma$ and μ?

→ Yes, since $\mu = 35$ and $\mu - 2\sigma = 35 - 2(8) = 19$.

FIGURE 7-5

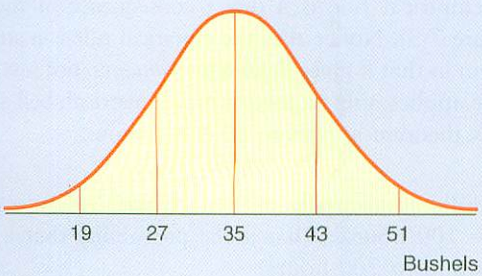
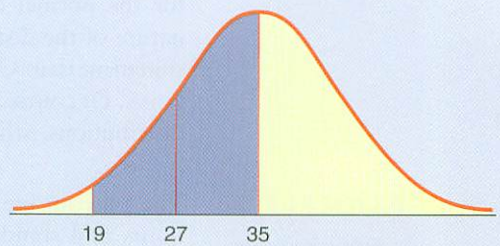


FIGURE 7-6 Completion of Figure 7-5

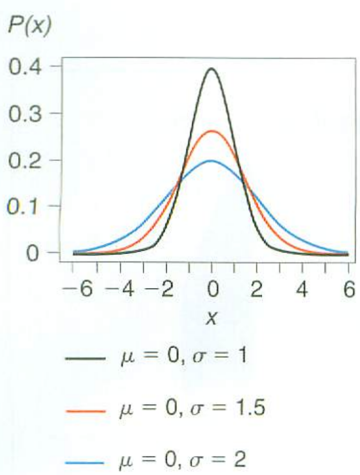


(c) Use Figure 7-3 to find the percentage of area over the interval between 19 and 35.

→ The area between the values $\mu - 2\sigma$ and μ is 47.5% of the total area.

(d) What is the probability that the yield will be between 19 and 35 bushels per acre?

→ The area is 47.5% of the total area, which is 1. Therefore, the probability is 0.475 that the yield will be between 19 and 35 bushels.



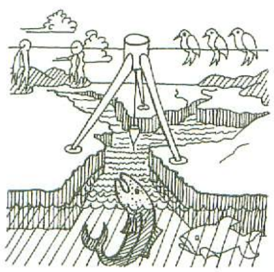
TECH NOTE We can graph normal distributions using the TI-84Plus and TI-83Plus calculators, Excel, and Minitab. In each technology, set the range of x values to between -3.5σ and 3.5σ. Then use the built-in normal density functions to generate the corresponding y values.

TI-84Plus/TI-83Plus Press the **Y=** key. Then, under **DISTR**, select **1:normalpdf** **(x,μ,σ)** and fill in the desired μ and σ values. Press the **WINDOW** key. Set **Xmin** to $\mu - 3\sigma$ and **Xmax** to $\mu + 3\sigma$. Finally, press the **ZOOM** key and select option **0:ZoomFit.**

Excel In one column, enter x values from -3.5σ to 3.5σ in increments of 0.2σ. In the next column, enter y values by using the menu choices **Paste function** (f_x) ➤ **Statistical** ➤ NORMDIST(x, μ, σ, false). Next, use the chart wizard, and select **XY(scatter)**. Choose the first picture with the dots connected and fill in the dialogue boxes.

Minitab In one column, enter x values from -3.5σ to 3.5σ in increments of 0.2σ. In the next column, enter y values by using the menu choices **Calc ➤ Probability Distribution ➤ Normal.** Fill in the dialogue box. Next, use the menu choices **Graph ➤ Plot.** Fill in the dialogue box. Under **Display,** select **Connect.**

VIEWPOINT



Nenana Ice Classic

The Nenana Ice Classic is a betting pool offering a large cash prize to the lucky winner who can guess the time, to the nearest minute, of the ice breakup on the Tanana River in the town of Nenana, Alaska. Official breakup time is defined as the time when the surging river dislodges a tripod on the ice. This breaks an attached line and stops a clock set to Yukon Standard Time. The event is so popular that the first state legislature of Alaska (1959) made the Nenana Ice Classic an official statewide lottery. Since 1918, the earliest breakup was April 20, 1940, at 3:27 P.M., and the latest recorded breakup was May 20, 1964, at 11:41 A.M. Want to make a statistical guess predicting when the ice will break up? Breakup times from 1918 to 1996 are recorded in *The Alaska Almanac,* published by Alaska Northwest Books, Anchorage.

SECTION 7.1 PROBLEMS

1. *General: Curves* Which, if any, of the curves in Figure 7-7 below look(s) like a normal curve? If a curve is not a normal curve, tell why.

2. *General: Normal Curves* Look at the normal curve in Figure 7-8 below, and find μ, $\mu + \sigma$, and σ.

FIGURE 7-7

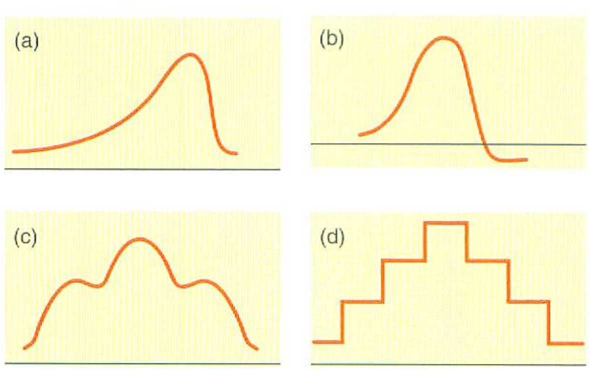
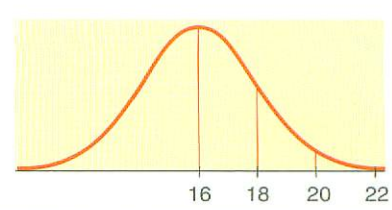


FIGURE 7-8



3. *General: Normal Curves* Look at the two normal curves in Figures 7-9 and 7-10. Which has the larger standard deviation? What is the mean of the curve in Figure 7-9? What is the mean of the curve in Figure 7-10?

FIGURE 7-9

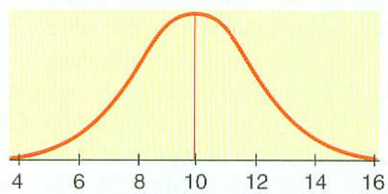
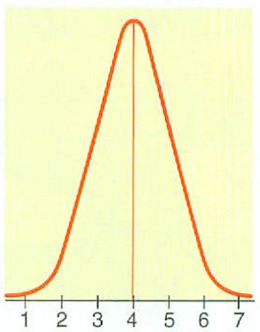


FIGURE 7-10



4. *General: Normal Curves* Sketch a normal curve
   (a) with mean 15 and standard deviation 2.
   (b) with mean 15 and standard deviation 3.
   (c) with mean 12 and standard deviation 2.
   (d) with mean 12 and standard deviation 3.
   (e) Consider two normal curves. If the first one has a larger mean than the second one, must it have a larger standard deviation as well? Explain your answer.

5. *General: Normal Curves* What percentage of the area under the normal curve lies
   (a) to the left of μ?
   (b) between $\mu - \sigma$ and $\mu + \sigma$?
   (c) between $\mu - 3\sigma$ and $\mu + 3\sigma$?

6. *General: Normal Curves* What percentage of the area under the normal curve lies
   (a) to the right of μ?
   (b) between $\mu - 2\sigma$ and $\mu + 2\sigma$?
   (c) to the right of $\mu + 3\sigma$?

7. *Distribution: Heights of Coeds* Assuming that the heights of college women are normally distributed, with mean 65 in. and standard deviation 2.5 in. (based on information from *Statistical Abstract of the United States*, 112th Edition), answer the following questions. (*Hint:* Use Problems 5 and 6 and Figure 7-3.)
   (a) What percentage of women are taller than 65 in.?
   (b) What percentage of women are shorter than 65 in.?
   (c) What percentage of women are between 62.5 in. and 67.5 in.?
   (d) What percentage of women are between 60 in. and 70 in.?

8. *Distribution: Rhode Island Red Chickens* The incubation time for Rhode Island Red chicks is normally distributed with a mean of 21 days and standard deviation of approximately 1 day (based on information from *World Book Encyclopedia*). Look at Figure 7-3 and answer the following questions. If 1000 eggs are being incubated, how many chicks do we expect will hatch
(a) in 19 to 23 days?
(b) in 20 to 22 days?
(c) in 21 days or fewer?
(d) in 18 to 24 days? (Assume all eggs eventually hatch.)
(*Note:* In this problem, let us agree to think of a single day or a succession of days as a continuous interval of time.)

9. *Archaeology: Tree Rings* At Burnt Mesa Pueblo, archaeological studies have used the method of tree-ring dating in an effort to determine when prehistoric people lived in the pueblo. Wood from several excavations gave a mean of (year) 1243 with a standard deviation of 36 years (*Bandelier Archaeological Excavation Project: Summer 1989 Excavations at Burnt Mesa Pueblo,* edited by Kohler, Washington State University Department of Anthropology). The distribution of dates was more or less mound-shaped and symmetrical about the mean. Use the empirical rule to
(a) estimate a range of years centered about the mean in which about 68% of the data (tree-ring dates) will be found.
(b) estimate a range of years centered about the mean in which about 95% of the data (tree-ring dates) will be found.
(c) estimate a range of years centered about the mean in which almost all the data (tree-ring dates) will be found.

10. *Vending Machine: Soft Drinks* A vending machine automatically pours soft drinks into cups. The amount of soft drink dispensed into a cup is normally distributed with a mean of 7.6 oz and standard deviation of 0.4 oz. Examine Figure 7-3 and answer the following questions.
(a) Estimate the probability that the machine will overflow an 8-oz cup.
(b) Estimate the probability that the machine will not overflow an 8-oz cup.
(c) The machine has just been loaded with 850 cups. How many of these do you expect will overflow when served?

11. *Pain Management: Laser Therapy* "Effect of Helium-Neon Laser Auriculotherapy on Experimental Pain Threshold" is the title of an article in the journal *Physical Therapy* (Vol. 70, No. 1, pp. 24–30). In this article, laser therapy was discussed as a useful alternative to drugs in pain management of chronically ill patients. To measure pain threshold, a machine was used that delivered low-voltage direct current to different parts of the body (wrist, neck, and back). The machine measured current in milliamperes (mA). The pretreatment experimental group in the study had an average threshold of pain (pain was first detectable) at $\mu = 3.15$ mA with standard deviation $\sigma = 1.45$ mA. Assume that the distribution of threshold pain so measured in milliamperes is symmetrical and more or less mound-shaped. Use the empirical rule to
(a) estimate a range of milliamperes centered about the mean in which about 68% of the experimental group will have a threshold of pain.
(b) estimate a range of milliamperes centered about the mean in which about 95% of the experimental group will have a threshold of pain.

7.2 Standard Units and Areas Under the Standard Normal Distribution

FOCUS POINTS

- Given μ and σ, convert raw data to z scores.
- Given μ and σ, convert z scores to raw data.
- Graph the standard normal distribution, and find areas under the standard normal curve.

z Scores and Raw Scores

Normal distributions vary from one another in two ways: The mean μ may be located anywhere on the x axis, and the bell shape may be more or less spread according to the size of the standard deviation σ. The differences among the normal distributions cause difficulties when we try to compute the area under the curve in a specified interval of x values and, hence, the probability that a measurement will fall into that interval.

It would be a futile task to try to set up a table of areas under the normal curve for each different μ and σ combination. We need a way to standardize the distributions so that we can use *one* table of areas for *all* normal distributions. We achieve this standardization by considering how many standard deviations a measurement lies from the mean. In this way, we can compare a value in one normal distribution with a value in another, different normal distribution. The next situation shows how this is done.

Suppose Tina and Jack are in two different sections of the same course. Each section is quite large, and the scores on the midterm exams of each section follow a normal distribution. In Tina's section, the average (mean) was 64 and her score was 74. In Jack's section, the mean was 72 and his score was 82. Both Tina and Jack were pleased that their scores were each 10 points above the average of each respective section. However, the fact that each person's score was 10 points above average does not really tell us how each person did *with respect to the other students in the section*. In Figure 7-11, we see the normal distribution of grades for each section.

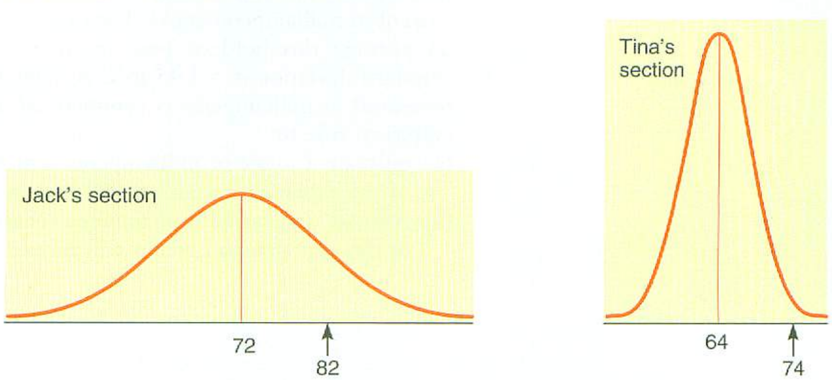
Tina's 74 was higher than most of the other scores in her section, while Jack's 82 is only an upper-middle score in his section. Tina's score is far better with respect to her class than Jack's score with respect to his class.

Standard score

The preceding situation demonstrates that it is not sufficient to know the difference between a measurement (x value) and the mean of a distribution. We need also to consider the spread of the curve, or the standard deviation. What we really want to know is the number of standard deviations between a measurement and the mean. This "distance" takes both μ and σ into account.

FIGURE 7-11

Distributions of Midterm Scores



We can use a simple formula to compute the number z of standard deviations between a measurement x and the mean μ of a normal distribution with standard deviation σ:

$$\left(\begin{array}{c}\text{Number of standard deviations}\\ \text{between the measurement and}\\ \text{the mean}\end{array}\right) = \left(\frac{\begin{array}{c}\text{Difference between the}\\ \text{measurement and the mean}\end{array}}{\text{Standard deviation}}\right)$$

z score

The z **value** or z **score** gives the number of standard deviations between the original measurement x and the mean μ of the x distribution.

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

TABLE 7-1
***x* Values and Corresponding *z* Values**

x Value in Original Distribution	Corresponding z Value or Standard Unit
$x = \mu$	$z = 0$
$x > \mu$	$z > 0$
$x < \mu$	$z < 0$

The mean is a special value of a distribution. Let's see what happens when we convert $x = \mu$ to a z value:

$$z = \frac{x - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

The mean of the original distribution is always zero, in standard units. This makes sense because the mean is zero standard deviations from itself.

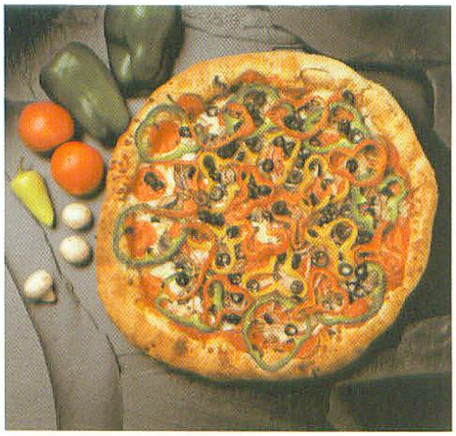
An x value in the original distribution that is *above* the mean μ has a corresponding z value that is *positive*. Again, this makes sense because a measurement above the mean would be a positive number of standard deviations from the mean. Likewise, an x value *below* the mean has a *negative* z value. (See Table 7-1.)

Note

Unless otherwise stated, in the remainder of the book we will take the word *average* to be either the sample arithmetic mean $\bar{x}$ or the population mean μ.

EXAMPLE 2
Standard score

A pizza parlor franchise specifies that the average (mean) amount of cheese on a large pizza should be 8 oz and the standard deviation only 0.5 oz. An inspector picks out a large pizza at random in one of the pizza parlors and finds that it is made with 6.9 oz of cheese. Assume that the amount of cheese on a pizza follows a normal distribution. If the amount of cheese is below the mean by more than *three* standard deviations, the parlor will be in danger of losing its franchise. (Remember, in a normal distribution we are unlikely to find measurements more than three standard deviations from the mean, since 99.7% of all measurements fall within three standard deviations of the mean.)



How many standard deviations from the mean is 6.9? Is the pizza parlor in danger of losing its franchise?

SOLUTION: Since we want to know the number of standard deviations from the mean, we want to convert 6.9 to standard z units.

$$z = \frac{x - \mu}{\sigma} = \frac{6.9 - 8}{0.5} = -2.20$$

Therefore, the amount of cheese on the selected pizza is only 2.20 standard deviations below the mean. Note that the fact that z is negative indicates that the amount of cheese was 2.20 standard deviations *below* the mean. The parlor will not lose its franchise based on this sample. ◊

GUIDED EXERCISE 3

Standard score

A student has computed that it takes an average (mean) of 17 minutes with a standard deviation of 3 minutes to drive from home, park the car, and walk to an early morning class.

(a) One day it took the student 21 minutes to get to class. How many standard deviations from the average is that? Is the z value positive or negative? Explain why it should be either positive or negative.

→ The number of standard deviations from the mean is given by the z value:

$$z = \frac{x - \mu}{\sigma} = \frac{21 - 17}{3} \approx 1.33$$

The z value is positive. We should expect a positive z value, since 21 minutes is *more* than the mean of 17.

(b) Another day it took only 12 minutes for the student to get to class. What is this measurement in standard units? Is the z value positive or negative? Why should it be positive or negative?

→ The measurement in standard units is

$$z = \frac{x - \mu}{\sigma} = \frac{12 - 17}{3} \approx -1.67$$

Here the z value is negative, as we should expect, because 12 minutes is less than the mean of 17 minutes.

(c) Another day it took 17 minutes for the student to go from home to class. What is the z value? Why should you expect this answer?

→ In this case, the z value is

$$z = \frac{x - \mu}{\sigma} = \frac{17 - 17}{3} = 0.00$$

We expect this result because 17 minutes is the mean, and the z value of the mean is always zero.

Raw score

We have seen how to convert from x measurements to standard units z. We can easily reverse the process to find the original *raw score* x if we know the mean and standard deviation of the original x distribution. Simply solve the z score formula for x.

Given an x distribution with mean μ and standard deviation σ, the raw score x corresponding to a z score is

$$x = z\sigma + \mu$$

EXAMPLE 3
Raw score

In Example 2, we talked about the amount of cheese required by a franchise for a large pizza. Again, the mean amount of cheese required is 8 oz with a standard deviation of 0.5 oz. The franchise specifies that the minimum amount of cheese for a large pizza is three standard deviations below the mean. A pizza parlor can lose its franchise if the amount of cheese on a large pizza is less than the specified minimum. What is the minimum amount of cheese that can be placed on a large pizza according to the franchise?

SOLUTION: Here we need to convert $z = -3$ to information about x oz of cheese. We use the formula

$$x = z\sigma + \mu = -3(0.5) + 8 = 6.5 \text{ oz}$$

The franchise will not approve a large pizza with less than 6.5 oz of cheese. ◆

In many testing situations, we hear the terms *raw score* and *z score*. The raw score is just the score in the original measuring units, and the z score is the score in standard units. Guided Exercise 4 illustrates these different units.

GUIDED EXERCISE 4

Raw score

Marulla's z score on a college entrance exam is 1.3. If the raw scores have a mean of 480 and a standard deviation of 70 points, what is her raw score?

➡ Here we are given z, σ, and μ. We need to find the raw score x corresponding to the z score 1.3.

$$\begin{aligned} x &= z\sigma + \mu \\ &= 1.3(70) + 480 \\ &= 571 \end{aligned}$$

Standard Normal Distribution

If the original distribution of x *values is normal,* then the corresponding z *values have a normal distribution as well.* The z distribution has a mean of 0 and a standard deviation of 1. The normal curve with these properties has a special name.

> The **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$ (Figure 7-12 on the next page).

Any normal distribution of x values can be converted to the standard normal distribution by converting all x values to their corresponding z values. Let's look at the graphic interpretation of this transformation in Figure 7-13 on the next page.

The resulting standard distribution will always have mean $\mu = 0$ and standard deviation $\sigma = 1$.

Areas Under the Standard Normal Curve

We have seen how to convert *any* normal distribution to the *standard* normal distribution. We can change any x value to a z value and back again. But what is the

FIGURE 7-12

The Standard Normal Distribution ($\mu = 0$, $\sigma = 1$)

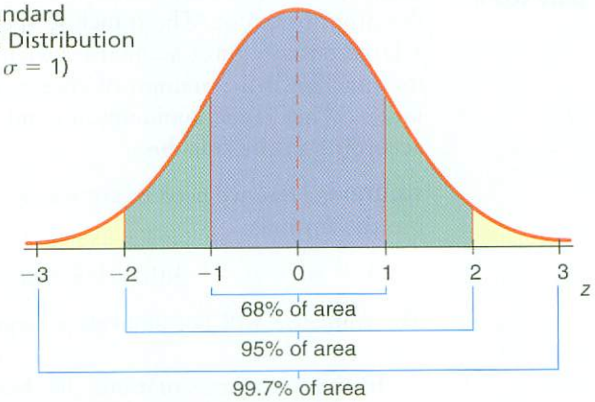
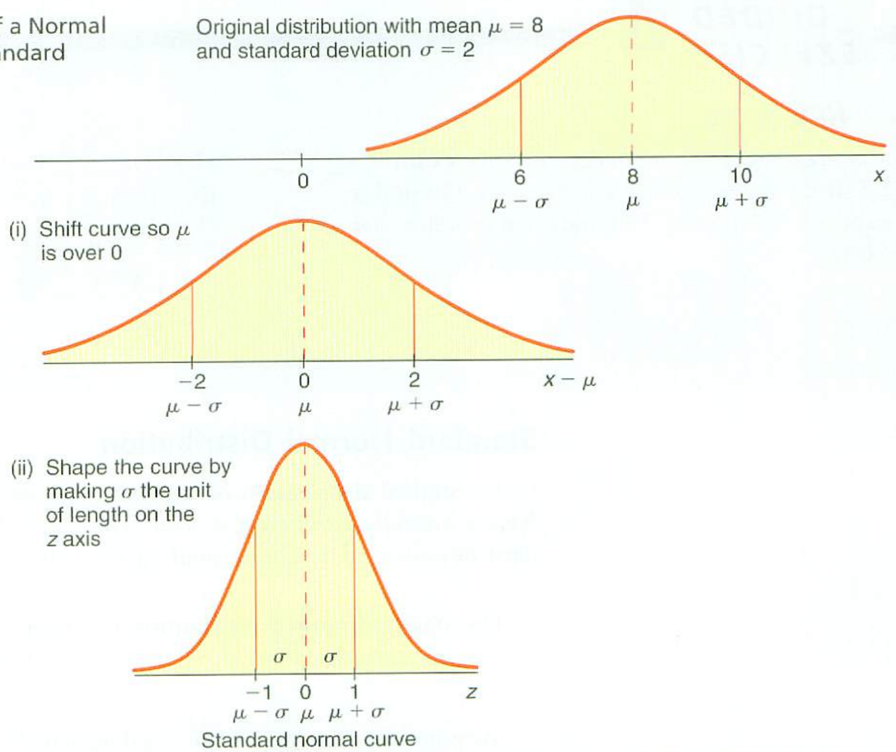


FIGURE 7-13

The Transformation of a Normal Distribution to the Standard Normal Distribution



advantage of all this work? The advantage is that there are extensive tables that show the *area under the standard normal curve* for almost any interval along the z axis. The areas are important because each area is equal to the *probability* that the measurement of an item selected at random falls in this interval. Thus, the *standard* normal distribution can be a tremendously helpful tool.

For instance, Sunshine Stereo guarantees their cassette decks for a period of 2 years. The company statistician has computed that the cassette deck life is normally distributed with a mean of 2.3 years and a standard deviation 0.4 year. What is the probability that a cassette deck will stop working during the guarantee period?

To answer questions of this type, we convert the given normal distribution to the standard normal distribution. Then we use a table, computer, or calculator to find the area over the interval in question and, hence, the probability that an item selected at random will fall into that interval. Before we can carry out this plan, though, let's practice using Table 3 of the Appendix to find areas under the standard normal curve.

Using a Standard Normal Distribution Table

Using a table to find areas and probabilities associated with the standard normal distribution is a fairly straightforward activity. However, it is important to first observe the range of z values for which areas are given. This range is usually depicted in a picture that accompanies the table.

Left-tail style table

In this text, *we will use the left-tail style table.* This style table gives cumulative areas to the left of a specified z. Determining other areas under the curve utilizes the fact that the area under the entire curve is 1. Taking advantage of the symmetry of the normal distribution is also useful. The procedures you learn for using the left-tail style normal distribution table apply directly to cumulative normal distribution areas found on calculators and in computer software packages such as Excel and Minitab.

EXAMPLE 4
Standard normal distribution table

Use Table 3 of the Appendix to find the described areas under the standard normal curve.

(a) Find the area under the standard normal curve to the left of $z = -1.00$.

SOLUTION: First, shade the area to be found on the standard normal distribution curve, as shown in Figure 7-14 on the next page. Notice that the z value we are using is negative. This means that we will look at the portion of Table 3 of the Appendix in which the z values are negative. In the upper-left corner of the table we see the letter z. The column under z gives us the units value and tenths for z. The other column headings indicate the hundredths value of z. Table entries give areas under the standard normal curve to the left of the listed z values. To find the area to the left of $z = -1.00$, we use the row headed by -1.0 and then move to the column headed by the hundredths position .00. This entry is shaded in Table 7-2. We see that the area is 0.1587.

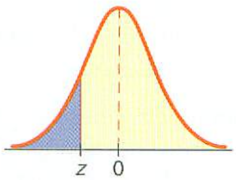


TABLE 7-2 Excerpt from Table 3 of the Appendix Showing Negative z Values

z	.00	.01	...	.07	.08	.09
−3.4	.0003	.0003	...	.0003	.0003	.0002
:						
−1.1	.1357	.1335	...	.1210	.1190	.1170
−1.0	.1587	.1562	...	.1423	.1401	.1379
−0.9	.1841	.1814	...	.1660	.1635	.1611
:						
−0.0	.5000	.4960	...	.4721	.4681	.4641

(b) Find the area to the left of $z = 1.18$, illustrated in Figure 7-15 on the next page.

FIGURE 7-14

Area to the Left of $z = -1.00$

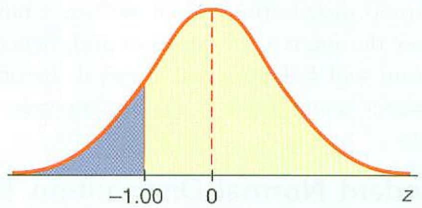
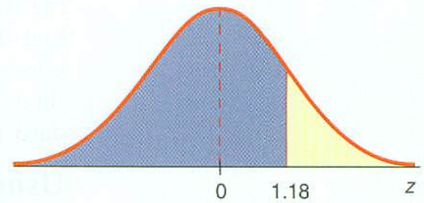


FIGURE 7-15

Area to the Left of $z = 1.18$



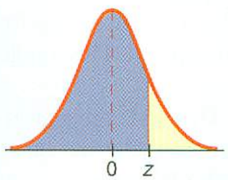


TABLE 7-3 Excerpt from Table 3 of the Appendix Showing Positive z Values

z	.00	.01	.02	...	.08	.09
0.0	.5000	.5040	.5080	...	.5319	.5359
⋮						
0.9	.8159	.8186	.8212	...	.8365	.8359
1.0	.8413	.8438	.8461	...	.8599	.8621
1.1	.8643	.8665	.8686	...	.8810	.8830
⋮						
3.4	.9997	.9997	.9997	...	.9997	.9998

SOLUTION: In this case, we are looking for an area to the left of a positive z value, so we look in the portion of Table 3 that shows positive z values. Again we first sketch the area to be found on a standard normal curve, as shown in Figure 7-15. Look in the row headed by 1.1 and move to the column headed by .08. The desired area is shaded (see Table 7-3). We see that the area to the left of 1.18 is 0.8810. ◇

GUIDED EXERCISE 5

Using the standard normal distribution table

Table 3, Areas of a Standard Normal Distribution, is located in the Appendix as well as in the endpapers of the text. Spend a little time studying the table, and then answer these questions.

(a) As z values increase, do the areas to the left of z increase?

→ Yes; as z values increase, we move to the right on the normal curve, and the areas increase.

(b) If a z value is negative, is the area to the left of z less than 0.5000?

→ Yes. Remember that a negative z value is on the left side of the standard normal distribution. The entire left half of the normal distribution has area 0.5, so any area to the left of $z = 0$ will be less than 0.5.

(c) If a z value is positive, is the area to the left of z greater than 0.5000?

→ Yes. Positive z values are on the right side of the standard normal distribution, and any area to the left of a positive z value includes the entire left half of the normal distribution.

Using Table 3 to find other areas

Table 3 gives areas under the standard normal distribution that are to the *left of a z value*. How do we find other areas under the standard normal curve?

PROCEDURE

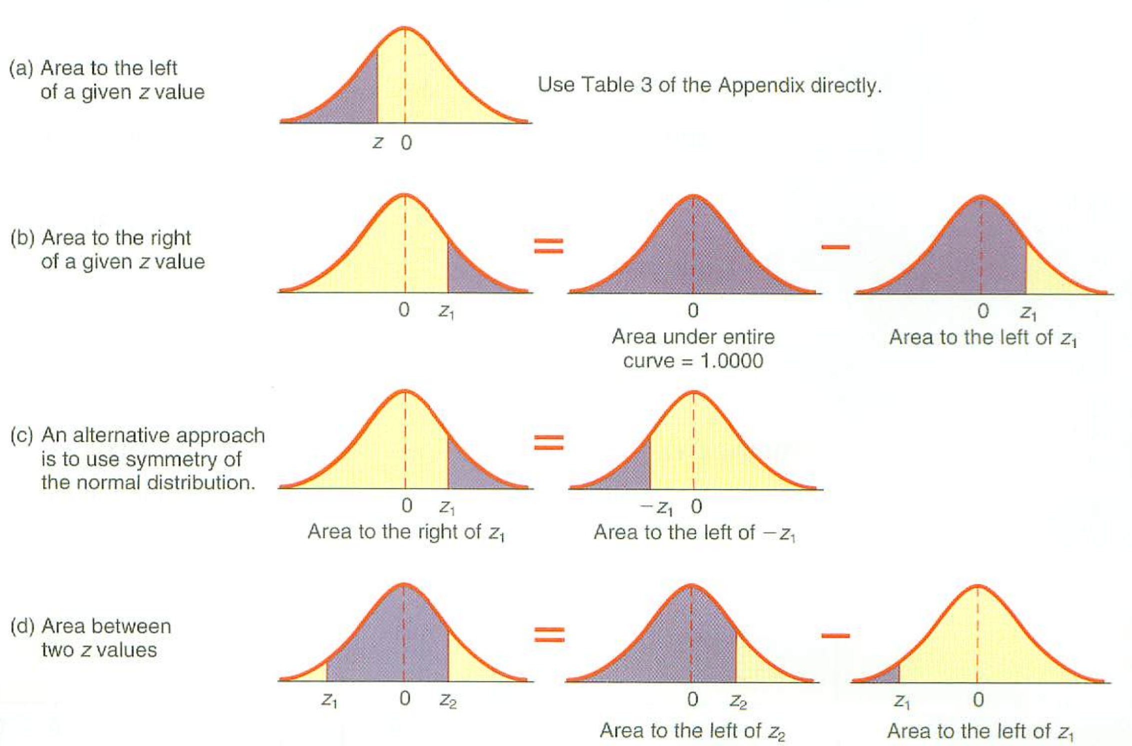
How to use a left-tail style standard normal distribution table

1. For areas to the left of a specified z value, use the table entry directly.
2. For areas to the right of a specified z value, look up the table entry for z and subtract the area from 1.
   Note: Another way to find the same area is to use the symmetry of the normal curve and look up the table entry for $-z$.
3. For areas between two z values z_1 and z_2 (where $z_2 > z_1$), *subtract* the table area for z_1 from the table area for z_2.

Figure 7-16 illustrates the procedure for using Table 3, Areas of a Standard Normal Distribution, to find any specified area under the standard normal distribution. Again, it is useful to sketch the area in question before you use Table 3.

◇ **COMMENT:** Notice that the z values shown in Table 3 of the Appendix are formatted to the hundredths position. It is convenient to *round or format z values to the hundredths position* before using the table. The areas are all given to four places after the decimal, so give your answers to four places after the decimal. ◇

FIGURE 7-16



◇ **COMMENT:** The smallest z value shown in Table 3 is -3.49, while the largest value is 3.49. These values are, respectively, far to the left and far to the right on the standard normal distribution, with very little area beyond either value. We will follow the common convention of treating any area to the left of a z value smaller than -3.49 as 0.000. Similarly, we will consider any area to the right of a z value greater than 3.49 as 0.000. We understand that there is some area in these extreme tails. However, these areas are each less than 0.0002. Now let's get real about this! Some very specialized applications, beyond the scope of this book, do need to measure areas and corresponding probabilities in these extreme tails. But in most practical applications, *we follow the convention of treating the areas in the extreme tails as zero.* ◇

Convention for using Table 3 of the Appendix

1. Treat any area to the left of a z value smaller than -3.49 as 0.000.
2. Treat any area to the left of a z value greater than 3.49 as 1.000.

EXAMPLE 5

Using table to find areas

Use Table 3 of the Appendix to find the specified areas.

(a) Find the area between $z = 1.00$ and $z = 2.70$.

SOLUTION: First, sketch a diagram showing the area (see Figure 7-17). Because we are finding the area between two z values, we subtract corresponding table entries.

$$\begin{aligned}(\text{Area between 1.00 and 2.70}) &= (\text{Area left of 2.70}) - (\text{Area left of 1.00}) \\ &= 0.9965 - 0.8413 = 0.1552\end{aligned}$$

(b) Find the area to the right of $z = 0.94$.

SOLUTION: First, sketch the area to be found (see Figure 7-18).

$$\begin{aligned}(\text{Area to right of 0.94}) &= (\text{Area under entire curve}) - (\text{Area to left of 0.94}) \\ &= 1.000 - 0.8264 = 0.1736\end{aligned}$$

Alternatively,

$$\begin{aligned}(\text{Area to right of 0.94}) &= (\text{Area to left of } -0.94) \\ &= 0.1736\end{aligned}$$

◇

FIGURE 7-17

Area from $z = 1.00$ to $z = 2.70$

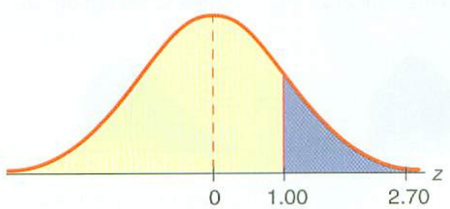
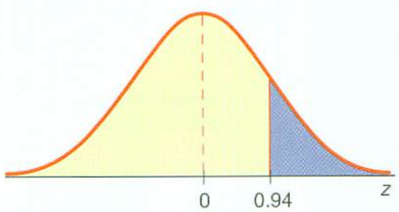


FIGURE 7-18

Area to the Right of $z = 0.94$



Probabilities associated with the standard normal distribution

We have practiced the skill of finding areas under the standard normal curve for various intervals along the z axis. This skill is important because *the probability* that z lies in an interval *is given by the area* under the standard normal curve above that interval.

Because the normal distribution is continuous, there is no area under the curve exactly over a specific z. Therefore, probabilities such as $P(z \geq z_1)$ are the same as $P(z > z_1)$. When dealing with probabilities or areas under a normal curve that are specified with inequalities, *strict inequality* symbols can be used *interchangeably* with *inequality-or-equal* symbols.

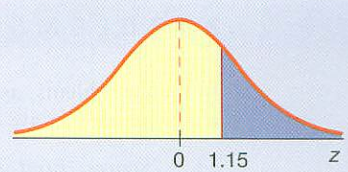
GUIDED EXERCISE 6

Probabilities associated with the standard normal distribution

Let z be a random variable with a standard normal distribution.

(a) $P(z \geq 1.15)$ refers to the probability that z values lie to the right of 1.15. Shade the corresponding area under the standard normal curve and find $P(z \geq 1.15)$.

FIGURE 7-19 Area to Be Found



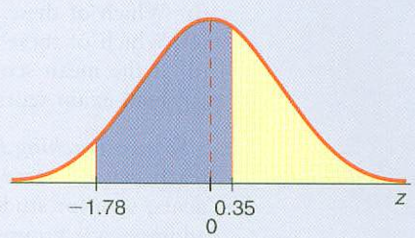
$$P(z \geq 1.15) = 1.000 - P(z \leq 1.15) = 1.000 - 0.8749 = 0.1251$$

Alternatively,

$$P(z \geq 1.15) = P(z \leq -1.15) = 0.1251$$

(b) Find $P(-1.78 \leq z \leq 0.35)$. First, sketch the area under the standard normal curve corresponding to the area.

FIGURE 7-20 Area to Be Found



$$P(-1.78 \leq z \leq 0.35) = P(z \leq 0.35) - P(z \leq -1.78)$$
$$= 0.6368 - 0.0375 = 0.5993$$

TECH NOTE The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all provide cumulative areas under any normal distribution, including the standard normal. The Tech Note of Section 7.3 shows examples.

VIEWPOINT

Mighty Oaks from Little Acorns Grow!

Just how big is that acorn? What if we compare it with other acorns? Is that oak tree taller than an average oak tree? How does it compare with other oak trees? What do you mean, this oak tree has a larger geographic range? Compared with what? Answers to questions such as these can be given only if we resort to *standardized statistical units.* Can you compare a single oak tree with an entire forest of oak trees? The answer is yes, if you use *standardized z scores.* For more information about sizes of acorns, oak trees, and geographic locations, visit the Brase/Brase statistics site at **http://math.college.hmco.com/students** and find the link to DASL, the Carnegie Mellon University Data and Story Library. From the DASL site, find Biology under Data Subjects, and select Acorns. Follow the links to Data Subjects, Biology, and Acorns.

SECTION 7.2 PROBLEMS

In these problems, assume that all the distributions are *normal.* In all problems in Chapter 7, *average* is always taken to be the arithmetic mean $\bar{x}$ or μ.

1. *z Scores: First Aid Course* A college Physical Education Department offered an Advanced First Aid course last semester. The scores on the comprehensive final exam were normally distributed, and the z scores for some of the students are shown below:

Robert, 1.10	Juan, 1.70	Susan, −2.00
Joel, 0.00	Jan, −0.80	Linda, 1.60

(a) Which of these students scored above the mean?
(b) Which of these students scored on the mean?
(c) Which of these students scored below the mean?
(d) If the mean score was $\mu = 150$ with standard deviation $\sigma = 20$, what was the final exam score for each student?

2. *z Scores: Teaching Duties* What do professors do with their time? They do research, teach classes, serve on academic committees, serve the student body (advise students, sponsor student clubs, attend student events), serve the community (consult, address civic groups), and a lot more! The specific answer depends on the individual professor and his or her special interests. Well, how much time does a professor spend on teaching activities? *The NEA Almanac of Higher Education,* published by the National Education Association, reports that the mean percentage of time professors spend on teaching activities is about $\mu = 51\%$ with standard deviation $\sigma = 25\%$. Find the standardized z values corresponding to the following professors' percentages of time allocated to teaching duties.
(a) Dr. Taylor, 45% (b) Mr. Patterson, 72% (c) Dr. Lee, 75%
(d) Ms Simms, 65% (e) Dr. Adams, 33% (f) Dr. Riley, 55%

3. *z Scores: Honolulu Temperatures* Data collected over a period of years show that the average daily temperature in Honolulu is $\mu = 73°F$ with standard deviation

$\sigma = 5°F$ (U.S. Department of Commerce, Environmental Data Service). Convert each of the following intervals in °F to an interval of z values.
(a) $53°F < x < 93°F$ (b) $x < 65°F$ (c) $78°F < x$

Convert the following intervals of z values to intervals in °F.
(d) $1.75 < z$ (e) $z < -1.90$ (f) $-1.80 < z < 1.65$

4. *z Scores: Fawns* Fawns between 1 and 5 months old in Mesa Verde National Park have a body weight that is approximately normally distributed with mean $\mu = 27.2$ kilograms and standard deviation $\sigma = 4.3$ kilograms (based on information from *The Mule Deer of Mesa Verde National Park,* by G. W. Mierau and J. L. Schmidt, Mesa Verde Museum Association). Let x be the weight of a fawn in kilograms. Convert each of the following x intervals to z intervals.
(a) $x < 30$ (b) $19 < x$ (c) $32 < x < 35$

Convert each of the following z intervals to x intervals.
(d) $-2.17 < z$ (e) $z < 1.28$ (f) $-1.99 < z < 1.44$
(g) If a fawn weighs 14 kilograms, would you say it is an unusually small animal? Explain using z values and Figure 7-12.
(h) If a fawn is unusually large, would you say that the z value for the weight of the fawn will be close to 0, -2, or 3? Explain.

5. *z Scores: Deer Population* The fall deer population in Mesa Verde National Park is approximately normally distributed with mean 4400 deer and standard deviation 620 deer (see reference in Problem 4). Let x be a random variable that represents the size of the deer population in Mesa Verde National Park in the fall of a given year. Convert each of the following x intervals to z intervals.
(a) $3300 < x$ (b) $x < 5400$ (c) $3500 < x < 5300$

Convert each of the following z intervals to x intervals.
(d) $-1.12 < z < 2.43$ (e) $z < 1.96$ (f) $2.58 < z$
(g) If the fall deer population were 2800 deer, would that be considered an unusually low number? If the fall population were 6300, would that be considered an unusually high population? Explain using z values and Figure 7-12.

6. *z Scores: White Blood Cell Count* Let x = white blood cell (WBC) count per cubic millimeter of whole blood. Then x has a distribution that is approximately normal with mean $\mu = 7500$ and standard deviation $\sigma = 1750$ (based on information from *Diagnostic Tests with Nursing Implications,* edited by S. Loeb, Springhouse Press). Convert each of the following x intervals to z intervals.
(a) $9000 < x$ (b) $x < 6000$ (c) $3500 < x < 4500$

Convert each of the following z intervals to x intervals.
(d) $z < 1.15$ (e) $2.19 < z$ (f) $0.25 < z < 1.25$
(g) If someone had a WBC count of 2500, would that be considered unusually high or low? Explain using z values and Figure 7-12.

7. *z Scores: Red Blood Cell Count* Let x = red blood cell (RBC) count in millions per cubic millimeter of whole blood. For healthy females x has an approximately normal distribution with mean $\mu = 4.8$ and standard deviation $\sigma = 0.3$. (See reference in Problem 6.) Convert each of the following x intervals from laboratory tests to z intervals.
(a) $4.5 < x$ (b) $x < 4.2$ (c) $4.0 < x < 5.5$

Convert each of the following z intervals to x intervals.
(d) $z < -1.44$ (e) $1.28 < z$ (f) $-2.25 < z < -1.00$
(g) If a female had an RBC count of 5.9 or higher, would that be considered unusually high? Explain using z values and Figure 7-12.

8. *Normal Curve: Tree Rings* Tree-ring dates were used extensively in archaeological studies at Burnt Mesa Pueblo (*Bandelier Archaeological Excavation Project: Summer 1989 Excavations at Burnt Mesa Pueblo*, edited by Kohler, Washington State University Department of Anthropology). At one site on the mesa, tree-ring dates (for many samples) gave a mean date of μ_1 = year 1272 with standard deviation $\sigma_1 =$ 35 years. At a second, removed site, the tree-ring dates gave a mean of μ_2 = year 1122 with standard deviation $\sigma_2 = 40$ years. Assume that both sites had dates that were approximately normally distributed. In the first area, an object was found and dated as x_1 = year 1250. In the second area, another object was found and dated as x_2 = year 1234.
   (a) Convert both x_1 and x_2 to z values, and locate both these values under the standard normal curve of Figure 7-12.
   (b) Which of these two items is the more unusual as an archaeological find in its location?

In Problems 9–28, sketch the areas under the standard normal curve over the indicated intervals, and find the specified areas.

9. To the right of $z = 0$.
10. To the left of $z = 0$.
11. To the left of $z = -1.32$.
12. To the left of $z = -0.47$.
13. To the left of $z = 0.45$.
14. To the left of $z = 0.72$.
15. To the right of $z = 1.52$.
16. To the right of $z = 0.15$.
17. To the right of $z = -1.22$.
18. To the right of $z = -2.17$.
19. Between $z = 0$ and $z = 3.18$.
20. Between $z = 0$ and $z = 2.92$.
21. Between $z = -2.01$ and $z = 0$.
22. Between $z = 0$ and $z = -1.93$.
23. Between $z = -2.18$ and $z = 1.34$.
24. Between $z = -1.40$ and $z = 2.03$.
25. Between $z = 0.32$ and $z = 1.92$.
26. Between $z = 1.42$ and $z = 2.17$.
27. Between $z = -2.42$ and $z = -1.77$.
28. Between $z = -1.98$ and $z = -0.03$.

In Problems 29–48, let z be a random variable with a standard normal distribution. Find the indicated probability, and shade the corresponding area under the standard normal curve.

29. $P(z \le 0)$.
30. $P(z \ge 0)$.
31. $P(z \le -0.13)$.
32. $P(z \le -2.15)$.
33. $P(z \le 1.20)$.
34. $P(z \le 3.20)$.
35. $P(z \ge 1.35)$.
36. $P(z \ge 2.17)$.
37. $P(z \ge -1.20)$.
38. $P(z \ge -1.50)$.
39. $P(-1.20 \le z \le 2.64)$.
40. $P(-2.20 \le z \le 1.04)$.
41. $P(-2.18 \le z \le -0.42)$.
42. $P(-1.78 \le z \le -1.23)$.
43. $P(0 \le z \le 1.62)$.
44. $P(0 \le z \le 0.54)$.
45. $P(-0.82 \le z \le 0)$.
46. $P(-2.37 \le z \le 0)$.
47. $P(-0.45 \le z \le 2.73)$.
48. $P(-0.73 \le z \le 3.12)$.

7.3 Areas Under Any Normal Curve

FOCUS POINTS

- Compute the probability of "standardized events."
- Find a z score from a given normal probability (inverse normal).
- Use the inverse normal to solve guarantee problems.

Normal Distribution Areas

In many applied situations, the original normal curve is not the standard normal curve. Generally, there will not be a table of areas available for the original normal curve. This does not mean that we cannot find the probability that a measurement x will fall into an interval from a to b. What we must do is *convert* the original measurements x, a, and b to z values.

PROCEDURE

How to work with normal distributions

To find areas and probabilities for a random variable x that follows a normal distribution with mean μ and standard deviation σ, convert x values to z values using the formula

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

Then use Table 3 of the Appendix to find corresponding areas and probabilities.

EXAMPLE 6 *Normal distribution probability*

Let x have a normal distribution with $\mu = 10$ and $\sigma = 2$. Find the probability that an x value selected at random from this distribution is between 11 and 14. In symbols, find $P(11 \le x \le 14)$.

SOLUTION: Since probabilities correspond to areas under the distribution curve, we want to find the area under the x curve above the interval from $x = 11$ to $x = 14$. To do so, we will convert the x values to standard z values (see Figure 7-21) and then use Table 3 of the Appendix to find the corresponding area under the standard curve.

We use the formula

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

to convert the given x interval to a z interval.

FIGURE 7-21

The Interval $11 \le x \le 14$ Corresponds to the Interval $0.50 \le z \le 2.00$

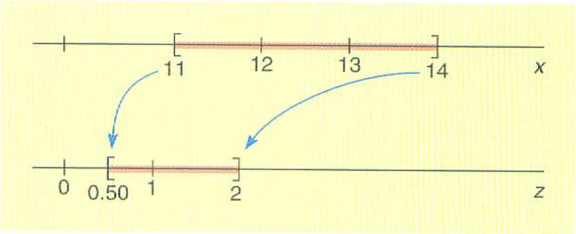
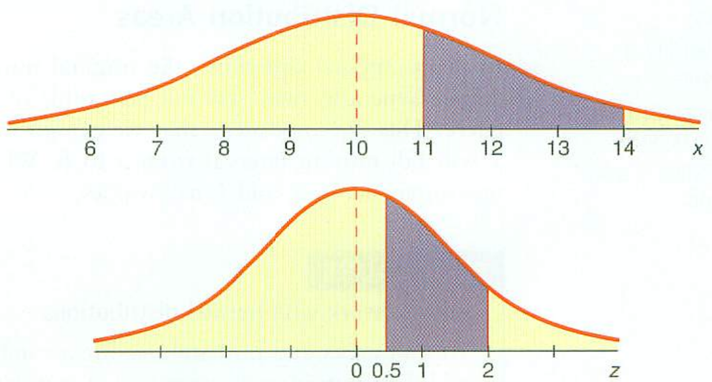


FIGURE 7-22

Corresponding Areas Under the x Curve and z Curve



$$z_1 = \frac{11 - 10}{2} = 0.50 \qquad \text{(Use } x = 11,\ \mu = 10,\ \sigma = 2.\text{)}$$

$$z_2 = \frac{14 - 10}{2} = 2.00 \qquad \text{(Use } x = 14,\ \mu = 10,\ \sigma = 2.\text{)}$$

The corresponding areas under the x and z curves are shown in Figure 7-22. From Figure 7-22 we see that

$$\begin{aligned} P(11 \le x \le 14) &= P(0.50 \le z \le 2.00) \\ &= P(z \le 2.00) - P(z \le 0.50) \\ &= 0.9772 - 0.6915 \qquad \text{(From Table 3 of the Appendix)} \\ &= 0.2857 \end{aligned}$$

The probability is 0.2857 that an x value selected at random from a normal distribution with mean 10 and standard deviation 2 lies between 11 and 14. ◆

GUIDED EXERCISE 7

Normal distribution probability

In Section 7.2, we talked about Sunshine Stereo cassette decks. The cassette deck life was normally distributed with a mean of 2.3 years and a standard deviation of 0.4 year. We wanted to know the probability that a cassette deck will break down during the guarantee period of 2 years.

(a) Let x represent the life of a cassette deck. The statement that the cassette deck breaks during the 2-year guarantee period means the life is less than 2 years, or $x \le 2$. Convert this to a statement about z.

➡ $z = \dfrac{x - \mu}{\sigma} = \dfrac{2 - 2.3}{0.4} = -0.75$

So $x \le 2$ means $z \le -0.75$.

Continued

GUIDED EXERCISE 7 continued

(b) Indicate the area to be found in Figure 7-23. Does this area correspond to the probability that $z \leq -0.75$?

➡ See Figure 7-24. Yes, the shaded area does correspond to the probability that $z \leq -0.75$.

FIGURE 7-23

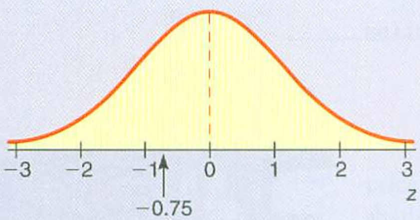
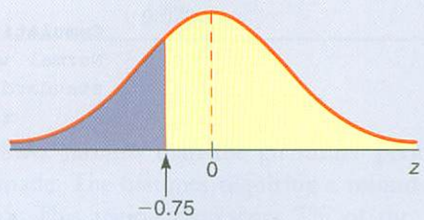


FIGURE 7-24 $z \leq -0.75$



(c) Use Table 3 of the Appendix to find $P(z \leq -0.75)$.

➡ 0.2266

(d) What is the probability that the cassette deck will break before the end of the guarantee period? [*Hint:* $P(x \leq 2) = P(z \leq -0.75)$.]

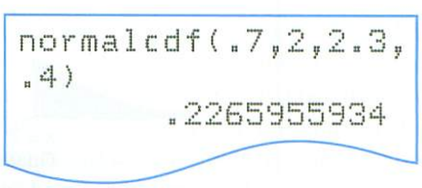
➡ The probability is

$$P(x \leq 2) = P(z \leq -0.75) = 0.2266$$

This means that the company will repair or replace about 23% of the cassette decks.

TECH NOTE The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all provide areas under any normal distribution. Excel and Minitab give the left-tail area to the left of a specified x value. The TI-84Plus/TI-83Plus has you specify an interval from a lower bound to an upper bound and provides the area under the normal curve for that interval. For example, to solve Guided Exercise 7 regarding the probability that a cassette deck will break during the guarantee period, we find $P(x \leq 2)$ for a normal distribution with $\mu = 2.3$ and $\sigma = 0.4$.

TI-84Plus/TI-83Plus Press the **DISTR** key, select **2:normalcdf (lower bound, upper bound, μ, σ)** and press **Enter**. Type in the specified values. For a left-tail area, use a lower bound setting at about four standard deviations below the mean. Likewise, for a right-tail area, use an upper bound setting about four standard deviations above the mean. For our example, use a lower bound of $\mu - 4\sigma = 2.3 - 4(0.4) = 0.7$.



Excel Select Paste Function (fx) ➤ Statistical ➤ NORMDIST. Fill in the dialogue box, using True for cumulative.

=	=NORMDIST(2,2.3,0.4, TRUE)	
C	D	E
0.226627		

Minitab Use the menu selection **Calc ➤ Probability Distribution ➤ Normal.** Fill in the dialogue box, marking cumulative.

```
Cumulative Distribution Function
Normal with mean = 2.3 and
standard deviation = 0.4
       x       P(X <= x)
     2.0          0.2266
```

Inverse Normal Distribution

Finding z or x, given a probability

Sometimes we need to find z or x values that correspond to a given area under the normal curve. This situation arises when we want to specify a guarantee period such that a given percentage of the total products produced by a company last at least as long as the duration of the guarantee period. In such cases, we use the standard normal distribution table "in reverse." When we look up an area and find the corresponding z value, we are using the *inverse normal probability distribution.*

EXAMPLE 7

Find x, given probability

Magic Video Games, Inc., sells an expensive video game console. Because the console is so expensive, the company wants to advertise an impressive guarantee for the life expectancy of its game console. The guarantee policy will refund the full purchase price if the console fails during the guarantee period. The research department has done tests which show that the mean life for the console is 30 months, with standard deviation of 4 months. The console life is normally distributed. How long can the guarantee period be if management does not want to refund the purchase price on more than 7% of the Magic Video Game Consoles?

SOLUTION: Let us look at the distribution of lifetimes for the computer game console, and shade the portion of the distribution in which the console lasts fewer months than the guarantee period. (See Figure 7-25.)

FIGURE 7-25

7% of the Consoles Have a Lifetime Less Than the Guarantee Period

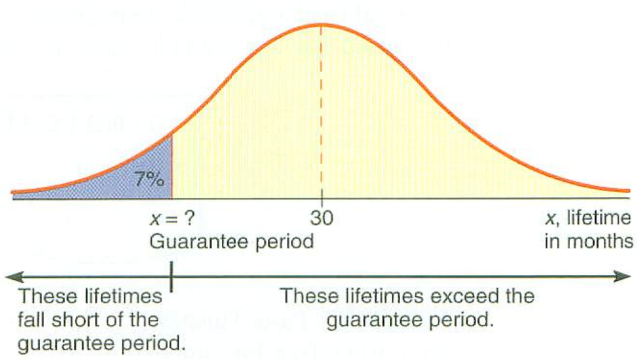


TABLE 7-4 Excerpt from Table 3 of the Appendix

z	.00	...	.07		.08	.09
:						
−1.4	.0808		.0708	↑	.0694	.0681
				0.0700		

If a game console lasts fewer months than the guarantee period, a full-price refund will have to be made. The lifetimes requiring a refund are in the shaded region of Figure 7-25. This region represents 7% of the total area under the curve.

We can use Table 3 of the Appendix to find the z value such that 7% of the total area under the *standard* normal curve lies to the left of the z value. Then we convert the z value to its corresponding x value to find the guarantee period.

We want to find the z value with 7% of the area under the standard normal curve to the left of z. Since we are given the area in a left tail, we can use Table 3 of the Appendix directly to find z. The area value is 0.0700. However, this area is not in our table, so we use the closest area, which is 0.0694, and the corresponding z value of $z = -1.48$ (see Table 7-4).

To translate this value back to an x value (in months), we use the formula

$$
\begin{aligned}
x &= z\sigma + \mu \\
&= -1.48(4) + 30 \qquad \text{(Use } \sigma = 4 \text{ months and } \mu = 30 \text{ months.)} \\
&= 24.08 \text{ months}
\end{aligned}
$$

The company can guarantee the Magic Video Game Console for $x = 24$ months. For this guarantee period, it expects to refund the purchase price of no more than 7% of the consoles. ◇

Example 7 had us find a z value corresponding to a given area to the left of z. What if the specified area is to the right of z or between $-z$ and z? Figure 7-26 on the next page shows us how to proceed.

◇ **COMMENT** When we use Table 3 of the Appendix to find a z value corresponding to a given area, we usually use the nearest area value rather than interpolating between values. However, when the area value given is exactly halfway between two area values of the table, we use the z value halfway between the z values of the corresponding table areas. Example 8 demonstrates this procedure. However, this interpolation convention is not always used, especially if the area is changing slowly, as it does in the tail ends of the distribution. *When the z value corresponding to an area is smaller than -2, the standard convention is to use the z value corresponding to the smaller area. Likewise, when the z value is larger than 2, the standard convention is to use the z value corresponding to the larger area.* We will see an example of this special case in Example 1 of Section 8.1. ◇

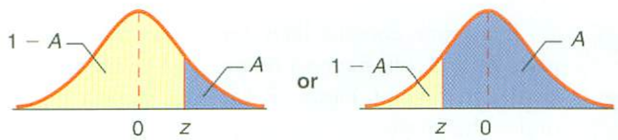
FIGURE 7-26

Inverse Normal: Use Table 3 of the Appendix to Find z Corresponding to a Given Area A $(0 < A < 1)$

(a) **Left-tail case:** The given area A is to the left of z.

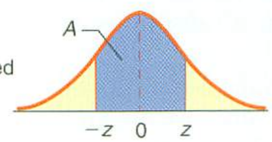
For the left-tail case, look up the number A in the body of the table and use the corresponding z value.

(b) **Right-tail case:** The given area A is to the right of z.



For the right-tail case, look up the number $1 - A$ in the body of the table and use the corresponding z value.

(c) **Center case:** The given area A is symmetric and centered above $z = 0$. Half of A lies to the left and half lies to the right of $z = 0$.



For the center case, look up the number $\frac{1-A}{2}$ in the body of the table and use the corresponding $\pm z$ value.

EXAMPLE 8
Find z

Find the z value such that 90% of the area under the standard normal curve lies between $-z$ and z.

SOLUTION: Sketch a picture showing the described area (see Figure 7-27).

We find the corresponding area in the left tail.

$$(\text{Area left of } -z) = \frac{1 - 0.9000}{2} = 0.0500$$

Looking at Table 7-5, we see that 0.0500 lies exactly between areas 0.0495 and 0.0505. The halfway value between $z = -1.65$ and $z = -1.64$ is $z = -1.645$. Therefore, we conclude that 90% of the area under the standard normal curve lies between the z values -1.645 and 1.645.

FIGURE 7-27

Area Between $-z$ and z Is 90%

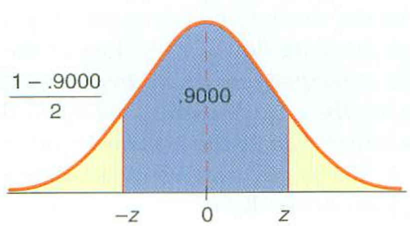


TABLE 7-5 Excerpt from Table 3 of the Appendix

z	...	.04		.05
:				
−1.6		.0505	↑	.0495
			0.0500	

◆

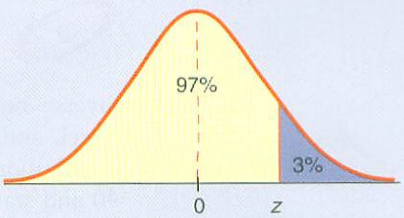
GUIDED EXERCISE 8

Find z

Find the z value such that 3% of the area under the standard normal curve lies to the right of z.

(a) Draw a sketch of the standard normal distribution showing the described area.

➡ **FIGURE 7-28** 3% of Total Area Lies to the Right of z



(b) Find the area to the left of z.

➡ The area to the left of $z = 1 - 0.0300 = 0.9700$.

(c) Look up the area in Table 7-6 and find the corresponding z.

➡ The closest area is 0.9699. This area is to the left of $z = 1.88$.

TABLE 7-6 Excerpt from Table 3 of the Appendix

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

(d) Suppose the time to complete a test is normally distributed with $\mu = 40$ minutes and $\sigma = 5$ minutes. After how many minutes can we expect all but about 3% of the tests to be completed?

➡ We are looking for an x value such that 3% of the normal distribution lies to the right of x. In part (c), we found that 3% of the standard normal curve lies to the right of $z = 1.88$. We convert $z = 1.88$ to an x value.

$$x = z\sigma + \mu$$
$$= 1.88(5) + 40 = 49.4 \text{ minutes}$$

All but about 3% of the tests will be complete after approximately 50 minutes.

Continued

GUIDED EXERCISE 8 continued

(e) Use Table 7-7 to find a z value such that 3% of the area under the standard normal curve lies to the left of z.

→ The closest area is 0.0301. This is the area to the left of $z = -1.88$.

TABLE 7-7 Excerpt from Table 3 of the Appendix

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294

(f) Compare the z value of part (c) with the z value of part (e). Is there any relationship between the z values?

→ One z value is the negative of the other. This result is expected because of the symmetry of the normal distribution.

TECH NOTE When we are given a z value and we find an area to the left of z, we are using a normal distribution function. When we are given an area to the left of z and we find the corresponding z, we are using an inverse normal distribution function. The TI-84Plus and TI-83Plus calculators, Excel, and Minitab all have inverse normal distribution functions for any normal distribution. For instance, to find an x value from a normal distribution with mean 40 and standard deviation 5 such that 97% of the area lies to the left of x, use the described instructions.

TI-84Plus/TI-83Plus Press the **DISTR** key and select **3:invNorm(area,μ,σ)**.

```
invNorm(.97,40,5)
          49.40396805
```

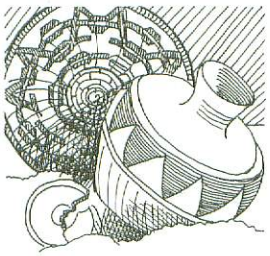
Excel Select **Paste Function** (f_x) ➤ **Statistical** ➤ **NORMINV**. Fill in the dialogue box.

=	=NORMINV(0.97,40,5)
C	D
49.40395	

Minitab Use the menu selection **Calc** ➤ **Probability Distribution** ➤ **Normal**. Fill in the dialogue box, marking Inverse Cumulative.

```
Inverse Cumulative Distribution Function
Normal with mean = 40.000 and
 standard deviation = 5.00000
P(X <= x)          x
   0.9700     49.4040
```

VIEWPOINT



Want to Be an Archaeologist?

Each year about 4500 students work with professional archaeologists in scientific research at the Crow Canyon Archaeological Center, Cortez, Colorado. In fact, Crow Canyon was included in *The Princeton Review Guide to America's Top 100 Internships.* The nonprofit, multidisciplinary program at Crow Canyon enables students and laypeople with little or no background to get started in archaeological research. The only requirement is that you be interested in Native American culture and history. By the way, a knowledge of introductory statistics could come in handy in this internship. For more information about the program, visit the Brase/Brase statistics site at **http://math.college.hmco.com/students** and find the link to Crow Canyon.

SECTION 7.3 PROBLEMS

In Problems 1–10, assume that x has a normal distribution, with the specified mean and standard deviation. Find the indicated probabilities.

1. $P(3 \le x \le 6)$; $\mu = 4$; $\sigma = 2$
2. $P(10 \le x \le 26)$; $\mu = 15$; $\sigma = 4$
3. $P(50 \le x \le 70)$; $\mu = 40$; $\sigma = 15$
4. $P(7 \le x \le 9)$; $\mu = 5$; $\sigma = 1.2$
5. $P(8 \le x \le 12)$; $\mu = 15$; $\sigma = 3.2$
6. $P(40 \le x \le 47)$; $\mu = 50$; $\sigma = 15$
7. $P(x \ge 30)$; $\mu = 20$; $\sigma = 3.4$
8. $P(x \ge 120)$; $\mu = 100$; $\sigma = 15$
9. $P(x \ge 90)$; $\mu = 100$; $\sigma = 15$
10. $P(x \ge 2)$; $\mu = 3$; $\sigma = 0.25$

In Problems 11–20, find the z value described and sketch the area described.

11. Find z such that 6% of the standard normal curve lies to the left of z.
12. Find z such that 5.2% of the standard normal curve lies to the left of z.
13. Find z such that 55% of the standard normal curve lies to the left of z.
14. Find z such that 97.5% of the standard normal curve lies to the left of z.
15. Find z such that 8% of the standard normal curve lies to the right of z.
16. Find z such that 5% of the standard normal curve lies to the right of z.
17. Find z such that 82% of the standard normal curve lies to the right of z.
18. Find z such that 95% of the standard normal curve lies to the right of z.
19. Find the z value such that 98% of the standard normal curve lies between $-z$ and z.
20. Find the z value such that 95% of the standard normal curve lies between $-z$ and z.
21. *Medical: Blood Glucose* A person's level of blood glucose and diabetes are closely related. Let x be a random variable measured in milligrams of glucose per deciliter (1/10 of a liter) of blood. After a 12-hour fast, the random variable x will have a distribution that is approximately normal with mean $\mu = 85$ and standard deviation $\sigma = 25$ (*Diagnostic Tests with Nursing Implications,* edited by S. Loeb, Springhouse

Press). *Note:* After 50 years of age, both the mean and standard deviation tend to increase. What is the probability that, for an adult (under 50 years old) after a 12-hour fast,

(a) x is more than 60?
(b) x is less than 110?
(c) x is between 60 and 110?
(d) x is greater than 140 (borderline diabetes starts at 140)?

22. *Medical: Blood Protoplasm* Porphyrin is a pigment in blood protoplasm and other body fluids that is significant in body energy and storage. Let x be a random variable that represents the number of milligrams of porphyrin per deciliter of blood. In healthy adults, x is approximately normally distributed with mean $\mu = 38$ and standard deviation $\sigma = 12$ (see reference in Problem 21). What is the probability that
(a) x is less than 60?
(b) x is greater than 16?
(c) x is between 16 and 60?
(d) x is more than 60? (This may indicate an infection, anemia, or another type of illness.)

23. *Education: SAT and ACT Scores* For a given population of high school seniors, the Scholastic Aptitude Test (SAT) in mathematics has a mean score of 500 with a standard deviation of 100. Another widely used test is the American College Testing (ACT) exam. The mathematics portion of the ACT has a mean of 18 and a standard deviation of 6. (Visit the Brase/Brase statistics site at **http://math.college.hmco.com/students** and find the link to the College Board.) Both SAT and ACT scores are normally distributed. What is the probability that a randomly selected high school senior's score on the mathematics part of the SAT will be
(a) more than 675? (b) less than 450? (c) between 450 and 675?

What is the probability that a randomly selected high school senior's score on the mathematics part of the ACT will be
(d) more than 28? (e) more than 12? (f) between 12 and 28?

24. *Inverse Normal Distribution: SAT and ACT Scores* Please refer to the SAT and ACT information from Problem 23.
(a) Suppose that an engineering school honors program will accept only high school seniors with a mathematics SAT or ACT score in the top 10%. What is the minimum SAT score in mathematics for this program? What is the minimum ACT score in mathematics for this program?
(b) Suppose that an engineering school will accept only high school seniors with a mathematics SAT or ACT score in the top 20%. What is the minimum SAT score in mathematics for this program? What is the minimum ACT score in mathematics for this program?
(c) Suppose that an engineering school will accept only high school seniors with a mathematics SAT or ACT score in the top 60%. What is the minimum SAT score in mathematics for this program? What is the minimum ACT score in mathematics for this program?

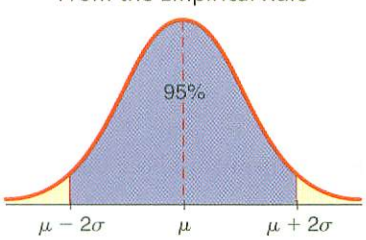
25. *Guarantee: Batteries* Quick Start Company makes 12-volt car batteries. After many years of product testing, the company knows that the average life of a Quick Start battery is normally distributed, with a mean of 45 months and a standard deviation of 8 months.
(a) If Quick Start guarantees a full refund on any battery that fails within the 36-month period after purchase, what percentage of its batteries will the company expect to replace?

(b) *Inverse Normal Distribution* If Quick Start does not want to make refunds for more than 10% of its batteries under the full-refund guarantee policy, for how long should the company guarantee the batteries (to the nearest month)?

26. *Guarantee: Watches* Accrotime is a manufacturer of quartz crystal watches. Accrotime researchers have shown that the watches have an average life of 28 months before certain electronic components deteriorate, causing the watch to become unreliable. The standard deviation of watch lifetimes is 5 months, and the distribution of lifetimes is normal.
   (a) If Accrotime guarantees a full refund on any defective watch for 2 years after purchase, what percentage of total production will the company expect to replace?
   (b) *Inverse Normal Distribution* If Accrotime does not want to make refunds on more than 12% of the watches it makes, how long should the guarantee period be (to the nearest month)?

27. *Expand Your Knowledge: Estimating the Standard Deviation* Consumer Reports gave information about the ages at which various household products are replaced. For example, color TVs are replaced at an average age of $\mu = 8$ years after purchase, and the (95% of data) range was from 5 to 11 years. Thus, the range was $11 - 5 = 6$ years. Let x be the age (in years) at which a color TV is replaced. Assume that x has a distribution that is approximately normal.
   (a) The empirical rule (Section 7.1) indicates that for a symmetrical and bell-shaped distribution, approximately 95% of the data lie within two standard deviations of the mean. Therefore, a 95% range of data values extending from $\mu - 2\sigma$ to $\mu + 2\sigma$ is often used for "commonly occurring" data values. Note that the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ is 4σ in length. This leads to a "rule of thumb" for estimating the standard deviation from a 95% range of data values.

From the Empirical Rule



Estimating the standard deviation

For a symmetric, bell-shaped distribution,

$$\text{standard deviation} \approx \frac{\text{range}}{4} \approx \frac{\text{high value} - \text{low value}}{4}$$

where it is estimated that about 95% of the commonly occurring data values fall into this range.

Use this "rule of thumb" to approximate the standard deviation of x values, where x is the age (in years) at which a color TV is replaced.
   (b) What is the probability that someone will keep a color TV for more than 5 years before replacement?
   (c) What is the probability that someone will keep a color TV for fewer than 10 years before replacement?
   (d) *Inverse Normal Distribution* Assume that the average life of a color TV is 8 years with a standard deviation of 1.5 years before it breaks. Suppose that a company guarantees color TVs and will replace a TV that breaks while under guarantee with a new one. However, the company does not want to replace more than 10% of the TVs under guarantee. For how long should the guarantee be made (rounded to the nearest tenth of a year)?

28. *Estimating the Standard Deviation: Veterinary Science* The resting heart rate for an adult horse should average about $\mu = 46$ beats per minute with a (95% of data)

range from 22 to 70 beats per minute, based on information from *The Merck Veterinary Manual* (a classic reference used in most veterinary colleges). Let x be a random variable that represents the resting heart rate for an adult horse. Assume that x has a distribution that is approximately normal.

(a) Estimate the standard deviation of the x distribution. *Hint:* See Problem 27.

(b) What is the probability that the heart rate is less than 25 beats per minute?

(c) What is the probability that the heart rate is greater than 60 beats per minute?

(d) What is the probability that the heart rate is between 25 and 60 beats per minute?

(e) ***Inverse Normal Distribution*** A horse whose resting heart rate is in the upper 10% of the probability distribution of heart rates may have a secondary infection or illness that needs to be treated. What is the heart rate corresponding to the upper 10% cutoff point of the probability distribution?

29\. ***Insurance: Satellites*** A relay microchip in a telecommunications satellite has a life expectancy that follows a normal distribution, with a mean of 90 months and a standard deviation of 3.7 months. When this computer-relay microchip malfunctions, the entire satellite is useless. A large London insurance company is going to insure the satellite for 50 million dollars. Assume that the only part of the satellite in question is the microchip. All other components will work indefinitely.

(a) ***Inverse Normal Distribution*** For how many months should the satellite be insured for the insurance company to be 99% confident that it will last beyond the insurance date?

(b) If the satellite is insured for 84 months, what is the probability that it will malfunction before the insurance coverage ends?

(c) If the satellite is insured for 84 months, what is the expected loss to the insurance company?

(d) If the insurance company charges \$3 million for 84 months of insurance, how much profit does the company expect to make?

7.4 Sampling Distributions

FOCUS POINTS

- ✓ Review such commonly used terms as random sample, relative frequency, parameter, statistic, and sampling distribution.
- ✓ From raw data, construct a relative frequency distribution for $\bar{x}$ values and compare the result to a theoretical sampling distribution.

Let us begin with some common statistical terms. Most of these have been discussed before, but this is a good time to review them.

From a statistical point of view, a *population* can be thought of as a set of measurements (or counts), either existing or conceptual. We discussed populations at some length in Chapter 1. A *sample* is a subset of measurements from the population. For our purposes, the most important samples are *random samples*, which were discussed in Section 1.2.

When we compute a descriptive measure such as an average, it makes a difference whether it was computed from a population or from a sample.

Statistic
Parameter

A **statistic** is a numerical descriptive measure of a *sample*.
A **parameter** is a numerical descriptive measure of a *population*.

It is important to notice that for a given population, a specified parameter is a fixed quantity. On the other hand, the value of a statistic might vary depending on which sample has been selected.

Some commonly used statistics and corresponding parameters

Measure	*Statistic*	*Parameter*
Mean	$\bar{x}$ (x bar)	μ (mu)
Variance	s^2	σ^2 (sigma squared)
Standard deviation	s	σ (sigma)
Proportion	$\hat{p}$ (p hat)	p

Often we do not have access to all the measurements of an entire population because of time, money, effort constraints, etc. So we must use measurements from a sample instead. In such cases, we will use a statistic (such as $\bar{x}$, s, or $\hat{p}$) to make *inferences* about corresponding population parameters (e.g., μ, σ, or p). The principal types of inferences we will make are the following.

Types of inferences

1. **Estimation:** In this type of inference, we estimate the *value* of a population parameter.
2. **Testing:** In this type of inference, we formulate a *decision* about the value of a population parameter.
3. **Regression:** In this type of inference, we make *predictions* or *forecasts* about the value of a statistical variable.

Sampling distribution

To evaluate the reliability of our inferences, we will need to know the probability distribution for the statistic we are using. Such a probability distribution is called a *sampling distribution.* Perhaps Example 9 below will help clarify this discussion.

A **sampling distribution** is a probability distribution of a sample statistic based on all possible simple random samples of the *same* size from the same population.

EXAMPLE 9

Sampling distribution for $\bar{x}$

Pinedale, Wisconsin, is a rural community with a children's fishing pond. Posted rules state that all fish under 6 inches must be returned to the pond, only children under 12 years old may fish, and a limit of five fish may be kept per day. Susan is a college student who was hired by the community last summer to make sure the rules were obeyed and to see that the children were safe from accidents. The pond contains only rainbow trout and has been well stocked for many years. Each child has no difficulty catching his or her limit of five trout.

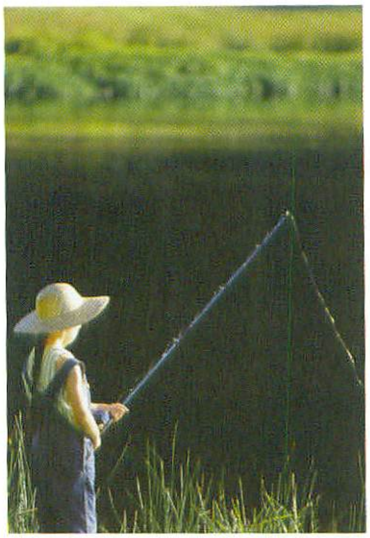
As a project for her biometrics class, Susan kept a record of the lengths (to the nearest inch) of all trout caught last summer. Hundreds of children visited the pond and caught their limit of five trout, so Susan has a lot of data. To make Table 7-8 (on the next page), Susan selected 100 children at random and listed the lengths of each of the five trout caught by each child in the sample. Then, for each child, she listed the mean length of the five trout that child caught.

TABLE 7-8 Length Measurements of Trout Caught by a Random Sample of 100 Children at the Pinedale Children's Pond

Sample	Length (to nearest inch)					$\bar{x}$ = Sample Mean	Sample	Length (to nearest inch)					$\bar{x}$ = Sample Mean
1	11	10	10	12	11	10.8	51	9	10	12	10	9	10.0
2	11	11	9	9	9	9.8	52	7	11	10	11	10	9.8
3	12	9	10	11	10	10.4	53	9	11	9	11	12	10.4
4	11	10	13	11	8	10.6	54	12	9	8	10	11	10.0
5	10	10	13	11	12	11.2	55	8	11	10	9	10	9.6
6	12	7	10	9	11	9.8	56	10	10	9	9	13	10.2
7	7	10	13	10	10	10.0	57	9	8	10	10	12	9.8
8	10	9	9	9	10	9.4	58	10	11	9	8	9	9.4
9	10	10	11	12	8	10.2	59	10	8	9	10	12	9.8
10	10	11	10	7	9	9.4	60	11	9	9	11	11	10.2
11	12	11	11	11	13	11.6	61	11	10	11	10	11	10.6
12	10	11	10	12	13	11.2	62	12	10	10	9	11	10.4
13	11	10	10	9	11	10.2	63	10	10	9	11	7	9.4
14	10	10	13	8	11	10.4	64	11	11	12	10	11	11.0
15	9	11	9	10	10	9.8	65	10	10	11	10	9	10.0
16	13	9	11	12	10	11.0	66	8	9	10	11	11	9.8
17	8	9	7	10	11	9.0	67	9	11	11	9	8	9.6
18	12	12	8	12	12	11.2	68	10	9	10	9	11	9.8
19	10	8	9	10	10	9.4	69	9	9	11	11	11	10.2
20	10	11	10	10	10	10.2	70	13	11	11	9	11	11.0
21	11	10	11	9	12	10.6	71	12	10	8	8	9	9.4
22	9	12	9	10	9	9.8	72	13	7	12	9	10	10.2
23	8	11	10	11	10	10.0	73	9	10	9	8	9	9.0
24	9	12	10	9	11	10.2	74	11	11	10	9	10	10.2
25	9	9	8	9	10	9.0	75	9	11	14	9	11	10.8
26	11	11	12	11	11	11.2	76	14	10	11	12	12	11.8
27	10	10	10	11	13	10.8	77	8	12	10	10	9	9.8
28	8	7	9	10	8	8.4	78	8	10	13	9	8	9.6
29	11	11	8	10	11	10.2	79	11	11	11	13	10	11.2
30	8	11	11	9	12	10.2	80	12	10	11	12	9	10.8
31	11	9	12	10	10	10.4	81	10	9	10	10	13	10.4
32	10	11	10	11	12	10.8	82	11	10	9	9	12	10.2
33	12	11	8	8	11	10.0	83	11	11	10	10	10	10.4
34	8	10	10	9	10	9.4	84	11	10	11	9	9	10.0
35	10	10	10	10	11	10.2	85	10	11	10	9	7	9.4
36	10	8	10	11	13	10.4	86	7	11	10	9	11	9.6
37	11	10	11	11	10	10.6	87	10	11	10	10	10	10.2
38	7	13	9	12	11	10.4	88	9	8	11	10	12	10.0
39	11	11	8	11	11	10.4	89	14	9	12	10	9	10.8
40	11	10	11	12	9	10.6	90	9	12	9	10	10	10.0
41	11	10	9	11	12	10.6	91	10	10	8	6	11	9.0
42	11	13	10	12	9	11.0	92	8	9	11	9	10	9.4
43	10	9	11	10	11	10.2	93	8	10	9	9	11	9.4
44	10	9	11	10	9	9.8	94	12	11	12	13	10	11.6
45	12	11	9	11	12	11.0	95	11	11	9	9	9	9.8
46	13	9	11	8	8	9.8	96	8	12	8	11	10	9.8
47	10	11	11	11	10	10.6	97	13	11	11	12	8	11.0
48	9	9	10	11	11	10.0	98	10	11	8	10	11	10.0
49	10	9	9	10	10	9.6	99	13	10	7	11	9	10.0
50	10	10	6	9	10	9.0	100	9	9	10	12	12	10.4

TABLE 7-9 Frequency Table for 100 Values of $\bar{x}$

Class	Class Limits: Lower	Class Limits: Upper	f = Frequency	$f/100$ = Relative Frequency
1	8.39	8.76	1	0.01
2	8.77	9.14	5	0.05
3	9.15	9.52	10	0.10
4	9.53	9.90	19	0.19
5	9.91	10.28	27	0.27
6	10.29	10.66	18	0.18
7	10.67	11.04	12	0.12
8	11.05	11.42	5	0.05
9	11.43	11.80	3	0.03



Now let us turn our attention to the following question: What is the average (mean) length of a trout taken from the Pinedale children's pond last summer?

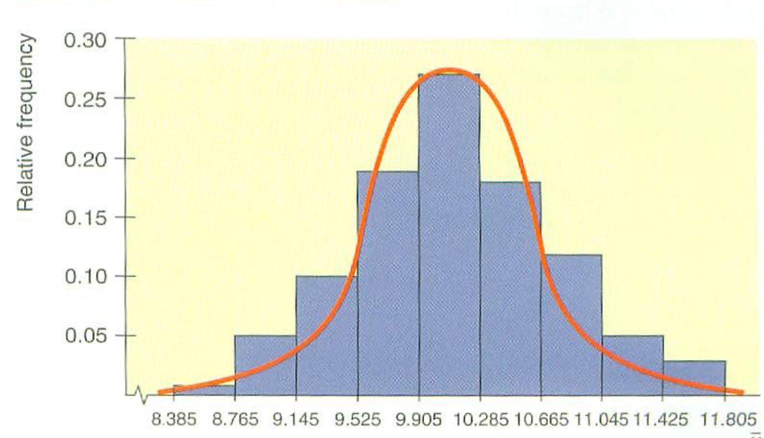
SOLUTION: We can get an idea of the average length by looking at the two $\bar{x}$ columns of Table 7-8. But just looking at 100 of the $\bar{x}$ values doesn't tell us much. Let's organize our $\bar{x}$ values into a frequency table. We used a class width of 0.38 to make Table 7-9.

Note: Techniques of Section 2.1 dictate a class width of 0.4. However, this choice results in the tenth class being beyond the data. Consequently, we shortened the class width slightly and also started the first class with a value slightly smaller than the smallest data value.

The far-right column of Table 7-9 contains relative frequencies $f/100$. Recall that the relative frequencies may be thought of as probabilities, so we effectively have a probability distribution. Because $\bar{x}$ represents the mean length of a trout (based on samples of five trout caught by each child), we estimate the probability of $\bar{x}$ falling into each class by using the relative frequencies. Figure 7-29 is a relative-frequency or probability distribution of the $\bar{x}$ values.

FIGURE 7-29

Estimates of Probabilities of $\bar{x}$ Values



The bars of Figure 7-29 represent our estimated probabilities of $\bar{x}$ values based on the data of Table 7-8. The bell-shaped curve represents the theoretical probability distribution that would be obtained if the number of children (i.e., number of $\bar{x}$ values) were much larger.

Figure 7-29 represents a *probability sampling distribution* for the sample mean $\bar{x}$ of trout lengths based on random samples of size 5. We see that the distribution is mound-shaped and even somewhat bell-shaped. Irregularities are due to the small number of samples used (only 100 sample means) and the rather small sample size (five trout per child). These irregularities would become less obvious and even disappear if the sample of children became much larger, if we used a larger number of classes in Figure 7-29, and if the number of trout used in each sample became larger. In fact, the curve would eventually become a perfect bell-shaped curve. We will discuss this property at some length in the next section, which introduces the *central limit theorem.* ◇

Let us summarize the information about sampling distributions in the following exercise.

GUIDED EXERCISE 9

Terminology

(a) What is a population parameter? Give an example.

→ A population parameter is a numerical descriptive measure of a population. Examples are μ, σ, and p. (There are many others.)

(b) What is a sample statistic? Give an example.

→ A sample statistic or a statistic is a numerical descriptive measure of a sample. Examples are $\bar{x}$, s, and $\hat{p}$.

(c) What is a sampling distribution?

→ A sampling distribution is a probability distribution for the sample statistic we are using.

(d) In Table 7-8, what makes up the members of the sample? What is the sample statistic corresponding to each sample? What is the sampling distribution? To which population parameter does this sampling distribution correspond?

→ There are 100 samples, each of which has five trout lengths. The first sample of five trout has lengths 11, 10, 10, 12, and 11. The sample statistic is the sample mean $\bar{x} = 10.8$. The sampling distribution is shown in Figure 7-29. This sampling distribution relates to the population mean μ of all lengths of trout taken from the Pinedale children's pond (i.e., trout over 6 inches long).

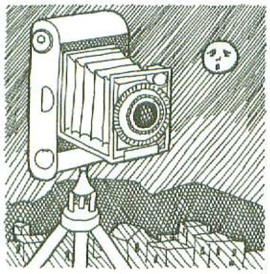
(e) Where will sampling distributions be used in our study of statistics?

→ Sampling distributions will be used for statistical inference. (Chapter 8 will concentrate on a method of inference called *estimation.* Chapter 9 will concentrate on a method of inference called *testing.*)

VIEWPOINT

"Chance Favors the Prepared Mind"

—Louis Pasteur



It also has been said that a discovery is nothing more than an accident that meets a prepared mind. Sampling can be one of the best forms of preparation. In fact, sampling may be the primary way we humans venture into the unknown. Probability sampling distributions can provide new information for the sociologist, scientist, or economist. In addition, ordinary human sampling of life can help writers and artists develop preferences, style, and insight. Ansel Adams became famous for photographing lyrical, unforgettable landscapes such as "Moonrise, Hernandez, New Mexico." Adams claimed that he was a strong believer in the quote by Pasteur. In fact, he claims that the Hernandez photograph was just such a favored chance happening that his prepared mind readily grasped. During his lifetime, Adams made over $25 million from sales and royalties on the Hernandez photograph.

SECTION 7.4 PROBLEMS

This is a good time to review several important concepts, some of which we have studied earlier. Please write out a careful but brief answer to each of the following questions.

1. What is a population? Give three examples.
2. What is a random sample from a population? (*Hint:* See Section 1.2.)
3. What is a population parameter? Give three examples.
4. What is a sample statistic? Give three examples.
5. What is the meaning of the term *statistical inference*? What types of inferences will we make about population parameters?
6. What is a sampling distribution?
7. How do frequency tables, relative frequencies, and histograms using relative frequencies help us understand sampling distributions?
8. How can relative frequencies be used to help us estimate probabilities occurring in sampling distributions?
9. Give an example of a specific sampling distribution we studied in this section. Outline other possible examples of sampling distributions from areas such as business administration, economics, finance, psychology, political science, sociology, biology, medical science, sports, engineering, chemistry, linguistics, and so on.

7.5 The Central Limit Theorem

FOCUS POINTS

- For a normal distribution, use μ and σ to construct the theoretical sampling distribution for the statistic $\bar{x}$.
- For large samples, use sample estimates to construct a good approximate sampling distribution for the statistic $\bar{x}$.
- Learn the statement and underlying meaning of the central limit theorem well enough to explain it to a friend who is intelligent, but (unfortunately) doesn't know much about statistics.

The $\bar{x}$ Distribution, Given x Is Normal

In Section 7.4, we began a study of the distribution of $\bar{x}$ values, where $\bar{x}$ was the (sample) mean length of five trout caught by children at the Pinedale children's fishing pond. Let's consider this example again in the light of a very important theorem of mathematical statistics.

◇ **THEOREM 7.1 for a Normal Probability Distribution** Let x be a random variable with a *normal distribution* whose mean is μ and whose standard deviation is σ. Let $\bar{x}$ be the sample mean corresponding to random samples of size n taken from the x distribution. Then the following are true:

(a) The $\bar{x}$ distribution is a *normal distribution.*

(b) The mean of the $\bar{x}$ distribution is μ.

(c) The standard deviation of the $\bar{x}$ distribution is $\sigma/\sqrt{n}$. ◇

We conclude from Theorem 7.1 that when x has a normal distribution, the $\bar{x}$ distribution will be normal *for any sample size n.* Furthermore, we can convert the $\bar{x}$ distribution to the standard normal z distribution using the following formulas.

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

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

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where n is the sample size,
μ is the mean of the x distribution, and
σ is the standard deviation of the x distribution.

Theorem 7.1 is a wonderful theorem! It states that the $\bar{x}$ distribution will be normal provided the x distribution is normal. The sample size n could be 2, 3, 4, or any (fixed) sample size we wish. Furthermore, the mean of the $\bar{x}$ distribution is μ (same as for the x distribution), but the standard deviation is $\sigma/\sqrt{n}$ (which is, of course, smaller than σ). The next example illustrates Theorem 7.1.

EXAMPLE 10

Probability regarding x; regarding $\bar{x}$

Suppose that a team of biologists has been studying the Pinedale children's fishing pond. Let x represent the length of a single trout taken at random from the pond. This group of biologists has determined that x has a normal distribution with mean $\mu = 10.2$ inches and standard deviation $\sigma = 1.4$ inches.

(a) What is the probability that a *single trout* taken at random from the pond is between 8 and 12 inches long?

SOLUTION: We use the methods of Section 7.3 with $\mu = 10.2$ and $\sigma = 1.4$ to get

$$z = \frac{x - \mu}{\sigma} = \frac{x - 10.2}{1.4}$$

Therefore,

$$\begin{aligned} P(8 < x < 12) &= P\left(\frac{8 - 10.2}{1.4} < z < \frac{12 - 10.2}{1.4}\right) \\ &= P(-1.57 < z < 1.29) \\ &= 0.9015 - 0.0582 = 0.8433 \end{aligned}$$

Therefore, the probability is about 0.8433 that a *single* trout taken at random is between 8 and 12 inches long.

(b) What is the probability that the *mean length* $\bar{x}$ of five trout taken at random is between 8 and 12 inches?

SOLUTION: If we let $\mu_{\bar{x}}$ represent the mean of the distribution, then Theorem 7.1, part (b), tells us that

$$\mu_{\bar{x}} = \mu = 10.2$$

If $\sigma_{\bar{x}}$ represents the standard deviation of the $\bar{x}$ distribution, then Theorem 7.1, part (c), tells us that

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.4/\sqrt{5} \approx 0.63$$

To create a standard z variable from $\bar{x}$, we subtract $\mu_{\bar{x}}$ and divide by $\sigma_{\bar{x}}$:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 10.2}{0.63}$$

To standardize the interval $8 < \bar{x} < 12$, we use 8 and then 12 in place of $\bar{x}$ in the preceding formula for z.

$$\begin{gathered} 8 < \bar{x} < 12 \\ \frac{8 - 10.2}{0.63} < z < \frac{12 - 10.2}{0.63} \\ -3.49 < z < 2.86 \end{gathered}$$

Theorem 7.1, part (a), tells us that $\bar{x}$ has a normal distribution. Therefore,

$$P(8 < \bar{x} < 12) = P(-3.49 < z < 2.86) = 0.9979 - 0.0002 = 0.9977$$

The probability is about 0.9977 that the mean length based on a sample size of 5 is between 8 and 12 inches.

(c) Looking at the results of parts (a) and (b), we see that the probabilities (0.8433 and 0.9977) are quite different. Why is this the case?

SOLUTION: According to Theorem 7.1, both x and $\bar{x}$ have a normal distribution, and both have the same mean of 10.2 inches. The difference is in the standard deviation for x and $\bar{x}$. The standard deviation of the x distribution is $\sigma = 1.4$. The standard deviation of the $\bar{x}$ distribution is

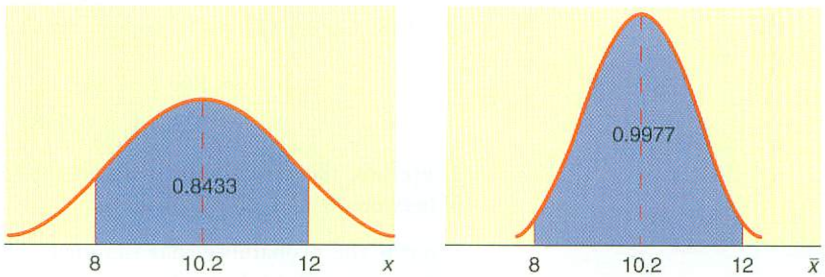
$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.4/\sqrt{5} \approx 0.63$$

FIGURE 7-30

General Shapes of the x and $\bar{x}$ Distributions

(a) The x distribution with $\mu = 10.2$ and $\sigma = 1.4$

(b) The $\bar{x}$ distribution with $\mu_{\bar{x}} = 10.2$ and $\sigma_{\bar{x}} = 0.63$ for samples of size $n = 5$



The standard deviation of $\bar{x}$ is less than half the standard deviation of x. Figure 7-30 shows the distributions of x and $\bar{x}$.

Looking at Figure 7-30(a) and (b), we see that both curves use the same scale on the horizontal axis. The means are the same, and the shaded area is above the interval from 8 to 12 on each graph. It becomes clear that the smaller standard deviation of the $\bar{x}$ distribution has the effect of gathering together much more of the total probability into the region over its mean. Therefore, the region from 8 to 12 has a much higher probability for the $\bar{x}$ distribution. ◇

Standard error of the mean

Theorem 7.1 describes the distribution of a particular statistic: namely, the distribution of sample means $\bar{x}$. The standard deviation of a statistic is referred to as the *standard error* of that statistic.

The **standard error** is the standard deviation of a sampling distribution. For the $\bar{x}$ sampling distribution,

$$\text{standard error} = \sigma_{\bar{x}} = \sigma/\sqrt{n}$$

Statistical software terminology

The expression *standard error* appears commonly on printouts and refers to the standard deviation of the sampling distribution being used. (In Minitab, the expression SE MEAN refers to the standard error of the mean.)

The $\bar{x}$ Distribution, Given x Follows Any Distribution

Central limit theorem

Theorem 7.1 gives complete information about the $\bar{x}$ distribution, provided the original x distribution is known to be normal. What happens if we don't have information about the shape of the original x distribution? The *central limit theorem* tells us what to expect.

◇ **THEOREM 7.2 The Central Limit Theorem for Any Probability Distribution** If x possesses *any* distribution with mean μ and standard deviation σ, then the sample mean $\bar{x}$ based on a random sample of size n will have a distribution that approaches the distribution of a normal random variable with mean μ and standard deviation $\sigma/\sqrt{n}$ as n increases without limit. ◇

The central limit theorem is indeed surprising! It says that x can have *any* distribution whatsoever, but as the sample size gets larger and larger, the distribution of $\bar{x}$ will approach a *normal* distribution. From this relation, we begin to appreciate the scope and significance of the normal distribution.

Large sample

In the central limit theorem, the degree to which the distribution of $\bar{x}$ values fits a normal distribution depends on both the selected value of n and the original distribution of x values. A natural question is: How large *should* the sample size be if we want to apply the central limit theorem? After a great deal of theoretical as well as empirical study, statisticians agree that if n is 30 or larger, the $\bar{x}$ distribution will appear to be normal and the central limit theorem will apply. However, this rule should not be applied blindly. If the x distribution is definitely not symmetrical about its mean, then the $\bar{x}$ distribution also will display a lack of symmetry. In such a case, a sample size larger than 30 may be required to get a reasonable approximation to the normal.

In practice, it is a good idea, when possible, to make a histogram of sample x values. If the histogram is approximately mound-shaped, and if it is more or less symmetrical, then we may be assured that, for all practical purposes, the $\bar{x}$ distribution will be well approximated by a normal distribution and the central limit theorem will apply when the sample size is 30 or larger. The main thing to remember is that in almost all practical applications, a sample size of 30 or more is adequate for the central limit theorem to hold. However, in a few rare applications, you may need a sample size larger than 30 to get reliable results.

Let's summarize this information for convenient reference: For almost all x distributions, if we use a random sample of size 30 or larger, the $\bar{x}$ distribution will be approximately normal, and the larger the sample size becomes, the closer the $\bar{x}$ distribution gets to the normal. Furthermore, we may convert the $\bar{x}$ distribution to a standard normal distribution using the following formulas.

Using the central limit theorem to convert the $\bar{x}$ distribution to the standard normal distribution

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

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

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where n is the sample size ($n \geq 30$),
μ is the mean of the x distribution, and
σ is the standard deviation of the x distribution.

Guided Exercise 10 shows how to standardize $\bar{x}$ when appropriate. Then Example 11 demonstrates the use of the central limit theorem in a decision-making process.

GUIDED EXERCISE 10

Central limit theorem

(a) Suppose x has a *normal* distribution with mean $\mu = 18$ and standard deviation $\sigma = 3$. If you draw random samples of size 5 from the x distribution and $\bar{x}$ represents the sample mean, what can you say about the $\bar{x}$ distribution? How could you standardize the $\bar{x}$ distribution?

→ Since the x distribution is given to be *normal*, the $\bar{x}$ distribution also will be normal even though the sample size is much less than 30. The mean is $\mu_{\bar{x}} = \mu = 18$. The standard deviation is

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 3/\sqrt{5} \approx 1.3$$

We could standardize $\bar{x}$ as follows:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 18}{1.3}$$

(b) Suppose you know that the x distribution has mean $\mu = 75$ and standard deviation $\sigma = 12$, but you have no information as to whether or not the x distribution is normal. If you draw samples of size 30 from the x distribution and $\bar{x}$ represents the sample mean, what can you say about the $\bar{x}$ distribution? How could you standardize the $\bar{x}$ distribution?

→ Since the sample size is large enough, the $\bar{x}$ distribution will be approximately a normal distribution. The mean of the $\bar{x}$ distribution is

$$\mu_{\bar{x}} = \mu = 75$$

The standard deviation of the distribution is

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 12/\sqrt{30} \approx 2.2$$

We could standardize $\bar{x}$ as follows:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 75}{2.2}$$

(c) Suppose you did not know that x had a normal distribution. Would you be justified in saying that the $\bar{x}$ distribution is approximately normal if the sample size were $n = 8$?

→ No, the sample size should be 30 or larger if we don't know that x has a normal distribution.

EXAMPLE 11
Central limit theorem

A certain strain of bacteria occurs in all raw milk. Let x be the bacteria count per milliliter of milk. The health department has found that if the milk is not contaminated, then x has a distribution that is more or less mound-shaped and symmetrical. The mean of the x distribution is $\mu = 2500$, and the standard deviation is $\sigma = 300$. In a large commercial dairy, the health inspector takes 42 random samples of the milk produced each day. At the end of the day, the bacteria count in each of the 42 samples is averaged to obtain the sample mean bacteria count $\bar{x}$.

(a) Assuming that the milk is not contaminated, what is the distribution of $\bar{x}$?

SOLUTION: The sample size is $n = 42$. Since this value exceeds 30, the central limit theorem applies, and we know that $\bar{x}$ will be approximately normal with mean and standard deviation

$$\mu_{\bar{x}} = \mu = 2500$$
$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 300/\sqrt{42} \approx 46.3$$

(b) Assuming the milk is not contaminated, what is the probability that the average bacteria count $\bar{x}$ for one day is between 2350 and 2650 bacteria per milliliter?

SOLUTION: We convert the interval

$$2350 \le \bar{x} \le 2650$$

to a corresponding interval on the standard z axis.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 2500}{46.3}$$

$$\bar{x} = 2350 \quad \text{converts to} \quad z = \frac{2350 - 2500}{46.3} \approx -3.24$$

$$\bar{x} = 2650 \quad \text{converts to} \quad z = \frac{2650 - 2500}{46.3} \approx 3.24$$

Therefore,

$$\begin{aligned} P(2350 \le \bar{x} \le 2650) &= P(-3.24 \le z \le 3.24) \\ &= 0.9994 - 0.0006 \\ &= 0.9988 \end{aligned}$$

The probability is 0.9988 that $\bar{x}$ is between 2350 and 2650.

(c) At the end of each day, the inspector must decide to accept or reject the accumulated milk that has been held in cold storage awaiting shipment. Suppose that the 42 samples taken by the inspector have a mean bacteria count $\bar{x}$ that is *not* between 2350 and 2650. If you were the inspector, what would be your comment on this situation?

SOLUTION: The probability that $\bar{x}$ is between 2350 and 2650 is very high. If the inspector finds that the average bacteria count for the 42 samples is not between 2350 and 2650, then it is reasonable to conclude that there is something wrong with the milk. If $\bar{x}$ is less than 2350, you might suspect someone added chemicals to the milk to artificially reduce the bacteria count. If $\bar{x}$ is above 2650, you might suspect some other kind of biologic contamination. ◇

PROCEDURE

How to find probabilities regarding $\bar{x}$

Given a probability distribution of x values where

n = sample size
μ = mean of the x distribution
σ = standard deviation of the x distribution

1. If the x distribution is *normal*, then the $\bar{x}$ distribution is *normal*.
2. Even if the x distribution is *not* normal, if the *sample size* $n \ge 30$, then, by the central limit theorem, the $\bar{x}$ distribution is *approximately normal*.
3. Convert $\bar{x}$ to z using the formula

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

4. Use the standard normal distribution to find the corresponding probability of events regarding $\bar{x}$.

GUIDED EXERCISE 11

Probability regarding $\bar{x}$

In mountain country, major highways sometimes use tunnels instead of long, winding roads over high passes. However, too many vehicles in a tunnel at the same time can cause a hazardous situation. Traffic engineers are studying a long tunnel in Colorado. If x represents the time for a vehicle to go through the tunnel, it is known that the x distribution has mean $\mu =$ 12.1 minutes and standard deviation $\sigma = 3.8$ minutes under ordinary traffic conditions. From a histogram of x values, it was found that the x distribution is mound-shaped with some symmetry about the mean.

Engineers have calculated that, *on average*, vehicles should spend from 11 to 13 minutes in the tunnel. If the time is less than 11 minutes, traffic is moving too fast for safe travel in the tunnel. If the time is more than 13 minutes, there is a problem of bad air (too much carbon monoxide and other pollutants).

Under ordinary conditions, there are about 50 vehicles in the tunnel at one time. What is the probability that the mean time for 50 vehicles to go through the tunnel will be from 11 to 13 minutes?

We will answer this question in steps.

(a) Let $\bar{x}$ represent the sample mean based on samples of size 50. Describe the $\bar{x}$ distribution.

→ From the central limit theorem, we expect the $\bar{x}$ distribution to be approximately normal with mean and standard deviation

$$\mu_{\bar{x}} = \mu = 12.1 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.8}{\sqrt{50}} \approx 0.54$$

(b) Find $P(11 < \bar{x} < 13)$.

→ We convert the interval

$$11 < \bar{x} < 13$$

to a standard z interval and use the standard normal probability table to find our answer. Since

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 12.1}{0.54}$$

$\bar{x} = 11$ converts to $z \approx \dfrac{11 - 12.1}{0.54} = -2.04$ and

$\bar{x} = 13$ converts to $z \approx \dfrac{13 - 12.1}{0.54} = 1.67$

Therefore,

$$\begin{aligned} P(11 < \bar{x} < 13) &= P(-2.04 < z < 1.67) \\ &= 0.9525 - 0.0207 \\ &= 0.9318 \end{aligned}$$

(c) Comment on your answer for part (b).

→ It seems that about 93% of the time there should be no safety hazard for average traffic flow.

VIEWPOINT

Chaos!

Is there a different side to random sampling? Can sampling be used as a weapon? According to *The Wall Street Journal*, the answer could be yes! The acronym for Create Havoc Around Our System is *CHAOS*. The Association of Flight Attendants (AFA) is a union that successfully used CHAOS against Alaska Airlines in 1994 as a negotiation tool. CHAOS involves a small sample of random strikes—a few flights at a time—instead of a mass walkout. The president of the AFA claims that by striking randomly, "we take control of the schedule." The entire schedule becomes unreliable, and that is something management cannot tolerate. In 1986, TWA flight attendants struck in a mass walkout, and all were permanently replaced! Using CHAOS, only a few jobs are put at risk, and these are usually not lost. It appears that random sampling *can* be used as a weapon.

SECTION 7.5 PROBLEMS

In these problems, the word *average* refers to the arithmetic mean $\bar{x}$ or μ, as appropriate.

1. *General* Suppose that x has a distribution with $\mu = 15$ and $\sigma = 14$.
   (a) If a random sample of size $n = 49$ is drawn, find $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$, and $P(15 \le \bar{x} \le 17)$.
   (b) If a random sample of size $n = 64$ is drawn, find $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$, and $P(15 \le \bar{x} \le 17)$.
   (c) Why should you expect the probability of part (b) to be higher than that of part (a)? (*Hint:* Consider the standard deviations in parts (a) and (b).)

2. *General* Suppose that x has a distribution with $\mu = 100$ and $\sigma = 48$.
   (a) If a random sample of size $n = 81$ is drawn, find $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$, and $P(92 \le \bar{x} \le 100)$.
   (b) If a random sample of size $n = 121$ is drawn, find $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$, and $P(92 \le \bar{x} \le 100)$.
   (c) Again, comment on the differences between the probabilities in parts (a) and (b). Why do you expect the differences?

3. *General* Suppose that x has a distribution with $\mu = 25$ and $\sigma = 3.5$.
   (a) If random samples of size $n = 9$ are selected, can we say anything about the $\bar{x}$ distribution of sample means?
   (b) If the original x distribution is normal, can we say anything about the $\bar{x}$ distribution from samples of size $n = 9$? Find $P(23 \le \bar{x} \le 26)$.

4. *General* Suppose that x has a distribution with $\mu = 72$ and $\sigma = 8$.
   (a) If random samples of size $n = 16$ are selected, can we say anything about the $\bar{x}$ distribution of sample means?
   (b) If the original x distribution is *normal*, can we say anything about the $\bar{x}$ distribution of random samples of size 16? Find $P(68 \le \bar{x} \le 73)$.

5. *Coal: Automatic Loader* Coal is carried from a mine in West Virginia to a power plant in New York in hopper cars on a long train. The automatic hopper car loader is set to put 75 tons of coal into each car. The actual weights of coal loaded into each car are *normally distributed* with mean $\mu = 75$ tons and standard deviation $\sigma = 0.8$ ton.
   (a) What is the probability that one car chosen at random will have less than 74.5 tons of coal?
   (b) What is the probability that 20 cars chosen at random will have a mean load weight $\bar{x}$ of less than 74.5 tons of coal?

(c) Suppose that the weight of coal in one car is less than 74.5 tons. Would that fact make you suspect that the loader had slipped out of adjustment? Suppose the weight of coal in 20 cars selected at random has an average $\bar{x}$ less than 74.5 tons. Would that fact make you suspect that the loader had slipped out of adjustment? Why?

6. *Vital Statistics: Heights of Men* The heights of 18-year-old men are approximately *normally distributed,* with mean 68 inches and standard deviation 3 inches (based on information from *Statistical Abstract of the United States,* 112th Edition).
   (a) What is the probability that an 18-year-old man selected at random is between 67 and 69 inches tall?
   (b) If a random sample of nine 18-year-old men is selected, what is the probability that the mean height $\bar{x}$ is between 67 and 69 inches?
   (c) Compare your answers for parts (a) and (b). Is the probability in part (b) much higher? Why would you expect this?

7. *Medical: Blood Glucose* Let x be a random variable that represents the level of glucose in the blood (milligrams per deciliter of blood) after a 12-hour fast. Assume that for people under 50 years old, x has a distribution that is approximately normal with mean $\mu = 85$ and estimated standard deviation $\sigma = 25$ (based on information from *Diagnostic Tests with Nursing Applications,* edited by S. Loeb, Springhouse). A test result $x < 40$ is an indication of severe excess insulin, and medication is usually prescribed.
   (a) What is the probability that, on a single test, $x < 40$?
   (b) Suppose a doctor uses the average $\bar{x}$ for two tests taken about a week apart. What can we say about the probability distribution of $\bar{x}$? *Hint:* See Theorem 7.1. What is the probability that $\bar{x} < 40$?
   (c) Repeat part (b) for $n = 3$ tests taken a week apart.
   (d) Repeat part (b) for $n = 5$ tests taken a week apart.
   (e) Compare your answers for parts (a), (b), (c), and (d). Did the probabilities decrease as n increased? Explain what this might suggest if you were a doctor or a nurse. If a patient had a test result of $\bar{x} < 40$ based on five tests, explain why either you are looking at an extremely rare event or (more likely) the person has a case of excess insulin.

8. *Medical: White Blood Cells* Let x be a random variable that represents white blood cell count per cubic milliliter of whole blood. Assume that x has a distribution that is approximately normal with mean $\mu = 7500$ and estimated standard deviation $\sigma = 1750$ (see reference in Problem 7). A test result of $x < 3500$ is an indication of leukopenia. This indicates bone marrow depression that may be the result of a viral infection.
   (a) What is the probability that, on a single test, x is less than 3500?
   (b) Suppose a doctor uses the average $\bar{x}$ for two tests taken about a week apart. What can we say about the probability distribution of $\bar{x}$? What is the probability of $\bar{x} < 3500$?
   (c) Repeat part (b) for $n = 3$ tests taken a week apart.
   (d) Compare your answers for parts (a), (b), and (c). How did the probabilities change as n increased? If a person had a test result of $\bar{x} < 3500$ based on three tests, what conclusion would you draw as a doctor or a nurse?

9. *Wildlife: Deer* Let x be a random variable that represents the weights in kilograms (kg) of healthy adult female deer (does) in December in Mesa Verde National Park. Then x has a distribution that is approximately normal with mean $\mu = 63.0$ kg and standard deviation $\sigma = 7.1$ kg (Source: *The Mule Deer of Mesa Verde National Park,* by G. W. Mierau and J. L. Schmidt, Mesa Verde Museum Association). Suppose a doe that weighs less than 54 kg is considered undernourished.
   (a) What is the probability that a single doe captured (weighed and released) at random in December is undernourished?

(b) If the park has about 2200 does, what number do you expect to be undernourished in December?
(c) To estimate the health of the December doe population, park rangers use the rule that the average weight of $n = 50$ does should be more than 60 kg. If the average weight is less than 60 kg, it is thought that the entire population of does might be undernourished. What is the probability that the average weight $\bar{x}$ for a random sample of 50 does is less than 60 kg (assume a healthy population)?
(d) Compute the probability that $\bar{x} < 64.2$ kg for 50 does (assume a healthy population). Suppose park rangers captured, weighed, and released 50 does in December, and the average weight was $\bar{x} = 64.2$ kg. Do you think the doe population is undernourished or not? Explain.

10. *Wildlife: Hummingbirds* *Selasphorus sasin* is the scientific name for what is commonly called "Allen's hummingbird." This beautiful hummingbird lives on the West Coast of the United States and is named for C. A. Allen (1841–1930), who studied these birds extensively. Let x be a random variable that represents the incubation time for Allen hummingbird eggs. Based on information from *The Hummingbird Book*, by Donald and Lillian Stokes (Little, Brown and Company), the x distribution has a mean of $\mu = 16$ days. Let us assume that the standard deviation is approximately $\sigma = 2$ days. The distribution of x values is more or less mound-shaped and symmetrical but not necessarily normal. Suppose that we have $n = 30$ eggs in an incubator. Let $\bar{x}$ be the average incubation time for these eggs.
(a) What can we say about the probability distribution of $\bar{x}$? Is it approximately normal? What are the mean and standard deviation?
(b) What is the probability that $\bar{x}$ is between 16 and 17 days?
(c) What is the probability that $\bar{x}$ is less than 15 days?

11. *Finance: Templeton Funds* Templeton World is a mutual fund that invests in both U.S. and foreign markets. Let x be a random variable that represents the monthly percentage return for the Templeton World fund. Based on information from the *Morningstar Guide to Mutual Funds* (available in most libraries), x has mean $\mu = 1.6\%$ and standard deviation $\sigma = 0.9\%$.
(a) Templeton World fund has over 250 stocks that combine together to give the overall monthly percentage return x. We can consider the monthly return of the stocks in the fund to be a sample from the population of monthly returns of all world stocks. Then we see that the overall monthly return x for Templeton World fund is itself an average return computed using all 250 stocks in the fund. Why would this indicate that x has an approximately normal distribution? Explain. *Hint:* See the discussion after Theorem 7.2.
(b) After 6 months, what is the probability that the *average* monthly percentage return $\bar{x}$ will be between 1% and 2%? *Hint:* See Theorem 7.1, and assume that x has a normal distribution as based on part (a).
(c) After 2 years, what is the probability that $\bar{x}$ will be between 1% and 2%?
(d) Compare your answers for parts (b) and (c). Did the probability increase as n (number of months) increased? Why would this happen?
(e) If after 2 years the average monthly percentage return $\bar{x}$ was less than 1%, would that tend to shake your confidence in the statement that $\mu = 1.6\%$? Might you suspect that μ has slipped below 1.6%? Explain.

12. *Focus Problem: Impulse Buying* Let x represent the dollar amount spent on supermarket impulse buying in a 10-minute (unplanned) shopping interval. Based on a *Denver Post* article, the mean of the x distribution is about \$20 and the estimated standard deviation is about \$7.
(a) Let us assume that x has a distribution that is approximately normal. What is the probability that x is between \$18 and \$22?

(b) Consider a random sample of $n = 100$ customers, each of whom has 10 minutes of unplanned shopping time in a supermarket. From the central limit theorem, what can you say about the probability distribution of $\bar{x}$, the average amount spent by these customers due to impulse buying? What are the mean and standard deviation of the $\bar{x}$ distribution? Is it necessary to make any assumption about the x distribution? Explain.

(c) What is the probability that $\bar{x}$ is between \$18 and \$22?

(d) In part (c) we used $\bar{x}$, the *average* amount spent, computed for 100 customers. In part (a) we used x, the amount spent by only *one* customer. The answers for parts (c) and (a) are very different. Why would this happen? In this example $\bar{x}$ is a much more predictable or reliable statistic than x. Consider that almost all marketing strategies and sales pitches are designed for the *average* customer and *not the individual* customer. How does the central limit theorem tell us that the average customer is much more predictable than the individual customer?

13. *General: Distribution of Sample Means*

(a) If we have a distribution of x values that is more or less mound-shaped and somewhat symmetrical, what is the sample size needed to claim that the distribution of sample means $\bar{x}$ from random samples of that size is approximately normal?

(b) If the original distribution of x values is known to be normal, do we need to make any restriction about sample size in order to claim that the distribution of sample means $\bar{x}$ taken from random samples of a given size is normal?

7.6 Normal Approximation to the Binomial Distribution

FOCUS POINTS

- ✓ State the necessary conditions for the normal approximation to the binomial distribution.
- ✓ Compute μ and σ for the normal approximation.
- ✓ Use the continuity correction to convert a range of r values to a corresponding range of normal x values.
- ✓ Convert the x values to a range of standardized z scores and find desired probabilities.

The probability that a new vaccine will protect adults from cholera is known to be 0.85. It is administered to 300 adults who must enter an area where the disease is prevalent. What is the probability that more than 280 of these adults will be protected from cholera by the vaccine?

This question falls into the category of a binomial experiment with the number of trials n equal to 300, the probability of success p equal to 0.85, and the number of successes r greater than 280. It is possible to use the formula for the binomial distribution to compute the probability that r is greater than 280. However, this approach would involve a number of tedious and long calculations. There is an easier way to do this problem, for under the conditions stated below, the normal distribution can be used to approximate the binomial distribution.

Large sample criteria $np > 5$ and $nq > 5$

Normal approximation to the binomial distribution

Consider a binomial distribution where

n = number of trials

r = number of successes

p = probability of success on a single trial

$q = 1 - p$ = probability of failure on a single trial

If $np > 5$ and $nq > 5$, then r has a binomial distribution that is approximated by a normal distribution with

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{npq}$$

Note: As n increases, the approximation becomes better.

Example 12 demonstrates that as n increases, the normal approximation to the binomial distribution improves.

EXAMPLE 12
Binomial distribution graphs

Graph the binomial distributions for which $p = 0.25$, $q = 0.75$, and the number of trials is first $n = 3$, then $n = 10$, then $n = 25$, and finally $n = 50$.

SOLUTION: The authors used a computer program to obtain the binomial distributions for the given values of p, q, and n. The results have been organized and graphed in Figures 7-31, 7-32, 7-33, and 7-34.

When $n = 3$, the outline of the histogram does not even begin to take the shape of a normal curve. But when $n = 10$, 25, or 50, it does begin to take a normal shape, indicated by the red curve. From a theoretical point of view, the histograms in Figures 7-32, 7-33, and 7-34 would have bars for all values of r from $r = 0$ to $r = n$. However, in the construction of these histograms, the bars of height less

FIGURE 7-31

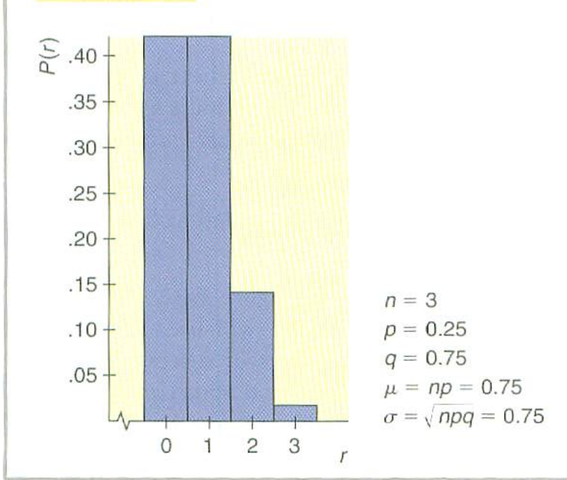


FIGURE 7-32

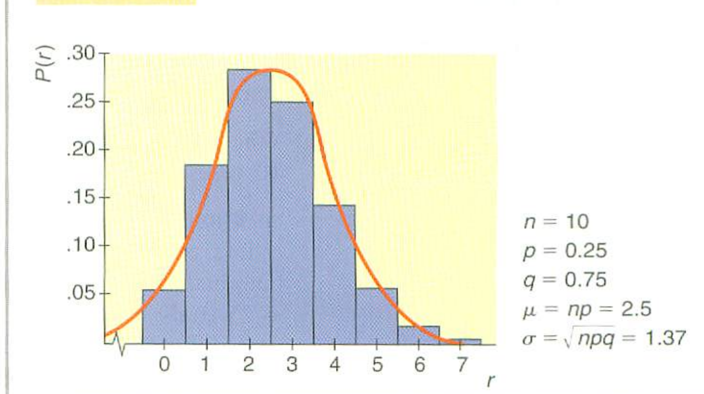


FIGURE 7-33

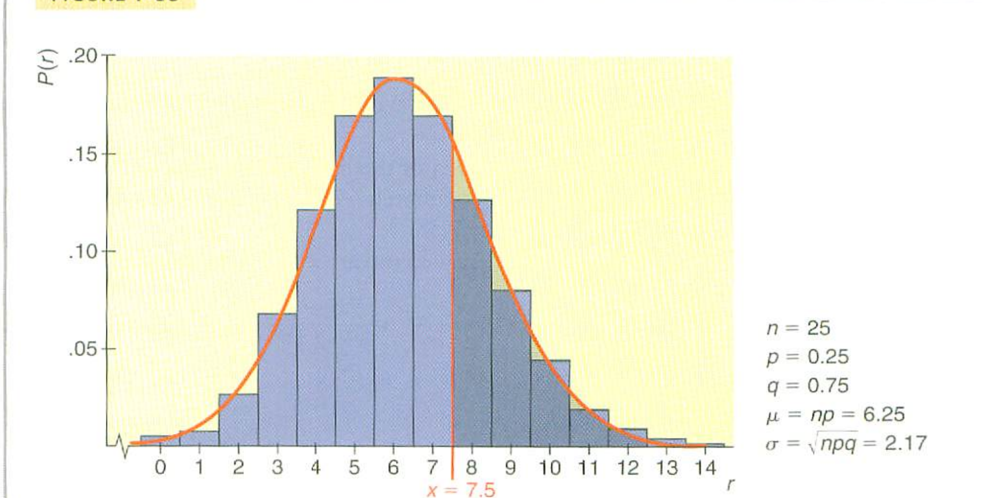
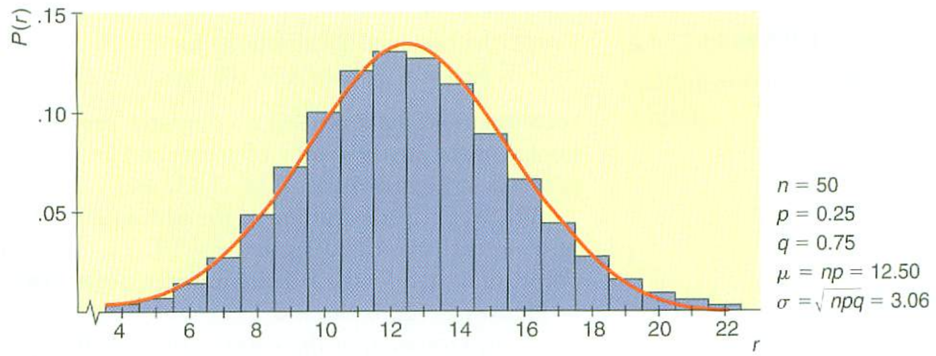


FIGURE 7-34



than 0.001 unit have been omitted—that is, in this example, probabilities less than 0.001 have been rounded to 0. ◆

EXAMPLE 13
Normal approximation

The owner of a new apartment building must install 25 water heaters. From past experience in other apartment buildings, she knows that Quick Hot is a good brand. A Quick Hot heater is guaranteed for 5 years only, but from her past experience, she knows that the probability it will last 10 years is 0.25.

(a) What is the probability that 8 or more of the 25 water heaters will last at least 10 years?

SOLUTION: In this example, $n = 25$ and $p = 0.25$, so Figure 7-33 (on the preceding page) represents the probability distribution we will use. Let r be the binomial random variable corresponding to the number of successes out of $n = 25$ trials. We want to find $P(r \geq 8)$ by using the normal approximation. This probability is represented graphically (Figure 7-33) by the area of the bar over 8 and all bars to the right of the bar over 8.

Let x be a normal random variable corresponding to a normal distribution with $\mu = np = 25(0.25) = 6.25$ and $\sigma = \sqrt{npq} = \sqrt{25(0.25)(0.75)} \approx 2.17$. This normal curve is represented by the red line in Figure 7-33. The area under the normal curve from $x = 7.5$ to the right is approximately the same as the area of the bars from the bar over $r = 8$ to the right. It is important to notice that we start with $x = 7.5$ because the bar over $r = 8$ really starts at $x = 7.5$.

The area of the bars and the area under the corresponding red (normal) curve are approximately equal, so we conclude that $P(r \geq 8)$ is approximately equal to $P(x \geq 7.5)$.

When we convert $x = 7.5$ to standard units, we get

$$z = \frac{x - \mu}{\sigma} = \frac{7.5 - 6.25}{2.17} \qquad \text{(use } \mu = 6.25 \text{ and } \sigma = 2.17\text{)}$$

$$\approx 0.58$$

The probability we want is

$$\begin{aligned} P(x \geq 7.5) &= P(z \geq 0.58) \\ &= 1 - P(z \leq 0.58) \\ &= 1 - 0.7190 \\ &= 0.2810 \end{aligned}$$

(b) How does this result compare with the result we can obtain by using the formula for the binomial probability distribution with $n = 25$ and $p = 0.25$?

SOLUTION: Using the binomial distribution function on the TI-84Plus/TI-83Plus model calculators, the authors computed that $P(r \geq 8) \approx 0.2735$. This means that the probability is approximately 0.27 that 8 or more water heaters will last at least 10 years.

(c) How do the results of parts (a) and (b) compare?

SOLUTION: The error of approximation is the difference between the approximate normal value (0.2810) and the binomial value (0.2735). The error is only $0.2810 - 0.2735 = 0.0075$, which is negligible for most practical purposes.

We knew in advance that the normal approximation to the binomial probability would be good, since $np = 25(0.25) = 6.25$ and $nq = 25(0.75) = 18.75$ are both greater than 5. These are the conditions that assure us that the normal approximation will be sufficiently close to the binomial probability for most practical purposes. ◆

Continuity correction: converting r values to x values

Remember that when using the normal distribution to approximate the binomial, we are computing the areas under bars. The bar over the discrete variable r extends from $r - 0.5$ to $r + 0.5$. This means that the corresponding continuous normal variable x extends from $r - 0.5$ to $r + 0.5$. Adjusting the values of discrete random variables to obtain a corresponding range for a continuous random variable is called making a *continuity correction.*

PROCEDURE

How to make the continuity correction

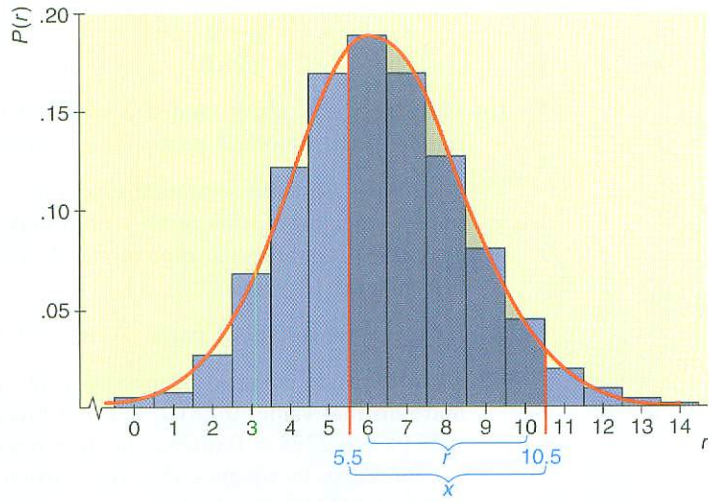
Convert the discrete random variable r (number of successes) to the continuous normal random variable x by doing the following:

1. If r is a **left-point** of an interval, subtract 0.5 to obtain the corresponding normal variable x; that is, $x = r - 0.5$.
2. If r is a **right-point** of an interval, add 0.5 to obtain the corresponding normal variable x: that is, $x = r + 0.5$.

For instance, $P(6 \leq r \leq 10)$, where r is a binomial random variable, is approximated by $P(5.5 \leq x \leq 10.5)$, where x is the corresponding normal random variable (see Figure 7-35 on the next page).

FIGURE 7-35

$P(6 \le r \le 10)$ Is Approximately Equal to $P(5.5 \le x \le 10.5)$



GUIDED EXERCISE 12

Continuity correction

From many years of observation, a biologist knows that the probability is only 0.65 that any given Arctic tern will survive the migration from its summer nesting area to its winter feeding grounds. A random sample of 500 Arctic terns were banded at their summer nesting area. Use the normal approximation to the binomial and the following steps to find the probability that between 310 and 340 of the banded Arctic terns will survive the migration. Let r be the number of surviving terns.

Arctic tern

(a) To approximate $P(310 \le r \le 340)$, we use the normal curve with μ = ______ and σ = ______.

➡ We use the normal curve with

$\mu = np = 500(0.65) = 325$ and

$\sigma = \sqrt{npq} = \sqrt{500(0.65)(0.35)} \approx 10.67$

(b) $P(310 \le r \le 340)$ is approximately equal to $P($______ $\le x \le$ ______$)$, where x is a variable from the normal distribution described in part (a).

➡ Since 310 is the left endpoint, we subtract 0.5, and since 340 is the right endpoint, we add 0.5. Consequently,

$P(310 \le r \le 340) \approx P(309.5 \le x \le 340.5)$

Continued

GUIDED EXERCISE 12 continued

(c) Convert the condition $309.5 \le x \le 340.5$ to a condition in standard units.

→ Since $\mu = 325$ and $\sigma \approx 10.67$, the condition $309.5 \le x \le 340.5$ becomes

$$\frac{309.5 - 325}{10.67} \le z \le \frac{340.5 - 325}{10.67} \quad \text{or}$$

$$-1.45 \le z \le 1.45$$

(d) $P(310 \le r \le 340) = P(309.5 \le x \le 340.5)$
$= P(-1.45 \le z \le 1.45)$
$=$ _______

→ $P(-1.45 \le z \le 1.45) = P(z \le 1.45) - P(z \le -1.45)$
$= 0.9265 - 0.0735$
$= 0.8530$

(e) Will the normal distribution make a good approximation to the binomial for this problem? Explain your answer.

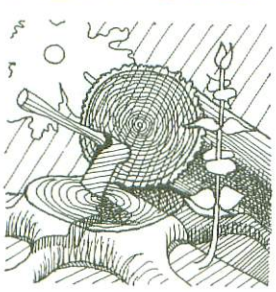
→ Since

$$np = 500(0.65) = 325 \quad \text{and}$$

$$nq = 500(0.35) = 175$$

are both greater than 5, the normal distribution will be a good approximation to the binomial.

VIEWPOINT



Sunspots, Tree Rings, and Statistics

Ancient Chinese astronomers recorded extreme sunspot activity with a peak around 1200 A.D. Mesa Verde tree rings in the period between 1276 and 1299 were unusually narrow, indicating a drought and/or a severe cold spell in the region at that time. A cooling trend could have narrowed the window of frost-free days below the approximately 80 days needed for cultivation of aboriginal corn and beans. Is this the reason the ancient Anasazi dwellings in Mesa Verde were abandoned? Is there a connection to the extreme sunspot activity? Much research and statistical work continues to be done on this topic.

Reference: *Prehistoric Astronomy in the Southwest*, by J. McKim Malville and C. Putnam, Department of Astronomy, University of Colorado.

SECTION 7.6 PROBLEMS

Note: When we say *between a* and *b*, we mean every value from *a* to *b*, *including a* and *b*. Due to rounding, your answers might vary slightly from answers given in the text.

1. *Health: Lead Contamination* More than a decade ago, high levels of lead in the blood put 88% of children at risk. A concerted effort was made to remove lead from the environment. Now, according to the *Third National Health and Nutrition Examination Survey* (*NHANES III*) conducted by the Centers for Disease Control, only 9% of children in the United States are at risk of high blood-lead levels.
   (a) In a random sample of 200 children taken more than a decade ago, what is the probability that 50 or more had high blood-lead levels?

(b) In a random sample of 200 children taken now, what is the probability that 50 or more have high blood-lead levels?

2. *Insurance: Claims* Do you try to *pad* an insurance claim to cover your deductible? About 40% of all U.S. adults will try to pad their insurance claims! (Source: *Are You Normal?* by Bernice Kanner, St. Martin's Press.) Suppose that you are the director of an insurance adjustment office. Your office has just received 128 insurance claims to be processed in the next few days. What is the probability that
   (a) half or more of the claims have been padded?
   (b) fewer than 45 of the claims have been padded?
   (c) from 40 to 64 of the claims have been padded?
   (d) more than 80 of the claims are *not* padded?

3. *Law Enforcement: Arrests* In Colorado Springs, a local newspaper ran a full page of nearly 100 mug shots of people the police wanted to arrest for serious crimes. Within 1 week, the police received enough information to locate and arrest about 17% of these "wanted" people (reported in *Rocky Mountain News*). If next month the newspaper runs a full page of 125 mug shots of fugitives, what is the probability that the police will receive enough information to locate and arrest (within 1 week)
   (a) at least 15 fugitives?
   (b) 28 or more fugitives?
   (c) between 15 and 28 fugitives?
   (d) In the solution to this problem, what is n? p? q? Does it appear that both np and nq are larger than 5? Why is this an important consideration?

4. *Novels: Romance* *USA Today* reported that 11% of all books sold are of the romance genre. If a local bookstore sells 316 books on a given day, what is the probability that
   (a) fewer than 40 are romances?
   (b) at least 25 are romances?
   (c) between 25 and 40 are romances?
   (d) In the solution to this problem, what is n? p? q? Does it appear that both np and nq are larger than 5? Why is this an important consideration?

5. *Longevity: 90th Birthday* It is estimated that 3.5% of the general population will live past their 90th birthday (*Statistical Abstract of the United States,* 112th Edition). In a graduating class of 753 high school seniors, what is the probability that
   (a) 15 or more will live beyond their 90th birthday?
   (b) 30 or more will live beyond their 90th birthday?
   (c) between 25 and 35 will live beyond their 90th birthday?
   (d) more than 40 will live beyond their 90th birthday?

6. *Fishing: Billfish* Ocean fishing for billfish is very popular in the Cozumel region of Mexico. In *World Record Game Fishes* (published by the International Game Fish Association), it was stated that in the Cozumel region about 44% of strikes (while trolling) resulted in a catch. Suppose that on a given day a fleet of fishing boats got a total of 24 strikes. What is the probability that the number of fish caught was
   (a) 12 or fewer?
   (b) 5 or more?
   (c) between 5 and 12?
   (d) In the solution to this problem, what is n? p? q? Does it appear that both np and nq are larger than 5? Why is this an important consideration?

7. *Grocery Stores: New Products* The *Denver Post* stated that 80% of all new products introduced in grocery stores fail (are taken off the market) within 2 years. If a

grocery store chain introduces 66 new products, what is the probability that within 2 years
(a) 47 or more fail?
(b) 58 or fewer fail?
(c) 15 or more succeed?
(d) fewer than 10 succeed?

8. *Crime: Murder* What are the chances that a person who is murdered actually knew the murderer? The answer to this question explains why a lot of police detective work begins with relatives and friends of the victim! About 64% of the people who are murdered actually knew the person who committed the murder (*Chances: Risk and Odds in Everyday Life,* by James Burke). Suppose that a detective file in New Orleans has 63 current unsolved murders. What is the probability that
(a) at least 35 of the victims knew their murderer?
(b) at most 48 of the victims knew their murderer?
(c) fewer than 30 victims did *not* know their murderer?
(d) more than 20 victims did *not* know their murderer?

9. *Private Investigation: Finding People* Old Friends Information Service is a California company that finds addresses for people who have lost track of each other. Old Friends claims to be 70% successful in reuniting people (*The Wall Street Journal*). In December, Old Friends had 430 requests for addresses of lost acquaintances. What is the probability that the number of addresses found was
(a) more than 280?
(b) at least 320?
(c) between 280 and 320?
(d) In the solution to this problem, what is n? p? q? Does it appear that both np and nq are larger than 5? Why is this an important consideration?

10. *Archaeology: Pottery* Santa Fe black-on-white is a style of pottery that occurs in about 61% of the pot shards found in the Bandelier National Monument area (*Bandelier Archaeological Excavation Project: Summer 1990 Excavations at Burnt Mesa Pueblo,* edited by Kohler, Washington State University). At one excavation site 8641 pot shards have been found that have not yet been cleaned and identified. What is the probability that
(a) fewer than 5200 are Santa Fe black-on-white?
(b) more than 5400 are Santa Fe black-on-white?
(c) between 5200 and 5400 are Santa Fe black-on-white?
(d) In the solution to this problem, what is n? p? q? Does it appear that both np and nq are larger than 5? Why is this an important consideration?

11. *Lawyers: Bar Exam* Over the years, it has been observed that of all the lawyers who take the state bar exam, only 57% pass (information from the National Conference on Bar Examiners, referenced in *The Book of Odds,* by Shook and Shook, Signet). Suppose that this year 850 lawyers are going to take the Ohio bar exam. What is the probability that
(a) 540 or more pass?
(b) 500 or fewer pass?
(c) between 485 and 525 pass?

12. *Ice Cream: Flavors* What's your favorite ice cream? For people who buy ice cream, the all-time favorite is still vanilla. About 25% of ice cream sales are vanilla. Chocolate accounts for only 9% of ice cream sales. (See reference in Problem 2.) Suppose that 175 customers go to a grocery store in Cheyenne, Wyoming, today to buy ice cream.
(a) What is the probability that 50 or more will buy vanilla?

(b) What is the probability that 12 or more will buy chocolate?

(c) A customer who buys ice cream is not limited to one container or one flavor. What is the probability that someone who is buying ice cream will buy chocolate or vanilla? *Hint:* Chocolate flavor and vanilla flavor are not mutually exclusive events. Assume that the choice to buy one flavor is independent of the choice to buy another flavor. Then use the multiplication rule for independent events together with the addition rule for events that are not mutually exclusive to compute the requested probability. (See Section 5.2.)

(d) What is the probability that between 50 and 60 customers buy chocolate or vanilla ice cream? *Hint:* Use the probability of success computed in part (c).

13. *Airline Flights: No-Shows* Based on long experience, an airline found that about 6% of the people making reservations on a flight from Miami to Denver do not show up for the flight. Suppose the airline overbooks this flight by selling 267 ticket reservations for an airplane with only 255 seats.

(a) What is the probability that a person holding a reservation will show up for the flight?

(b) Let $n = 267$ represent the number of ticket reservations. Let r represent the number of people with reservations who show up for the flight. Which expression represents the probability that a seat will be available for everyone who shows up holding a reservation?

$$P(255 \le r); \quad P(r \le 255); \quad P(r \le 267); \quad P(r = 255)$$

(c) Use the normal approximation to the binomial distribution and part (b) to answer the following question: What is the probability that a seat will be available for every person who shows up holding a reservation?

SUMMARY

In this chapter, we examined normal distributions, z scores, raw scores, area under normal curves, $\bar{x}$ sampling distributions, and normal approximations to binomial distributions.

The standard normal distribution has mean $\mu = 0$ and standard deviation $\sigma = 1$. To convert a normal distribution with mean μ and standard deviation σ to the standard normal distribution, we use the formula

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

Probabilities associated with the standard normal distribution are found in Table 3 of the Appendix.

Sampling distributions give us the basis for inferential statistics. By studying the distribution of sample statistics, we can learn about a population parameter.

The central limit theorem describes the sampling distribution of sample means $\bar{x}$ taken from samples of size n. It tells us that for increasing sample size n, the distribution of sample means approaches a normal distribution with mean $\mu_{\bar{x}} = \mu$

and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, where μ and σ are the respective mean and standard deviation of the original x distribution. Using the central limit theorem, we convert $\bar{x}$ values to standard z scores by the formula

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

We can use the normal distribution to approximate the binomial distribution if $np > 5$ and $nq > 5$, where n is the number of trials, p is the probability of success on a single trial, and $q = 1 - p$.

IMPORTANT WORDS & SYMBOLS

Section 7.1
Normal distributions
Normal curves
Upward cup and downward cup on normal curves
Symmetry of normal curves
Empirical rule

Section 7.2
z value or z score
Standard units
Standard normal distribution ($\mu = 0$ and $\sigma = 1$)
Raw score, x
Area under the standard normal curve

Section 7.3
Areas under any normal curve

Section 7.4
Population parameter
Statistic
Sampling distribution

Section 7.5
$\mu_{\bar{x}}$
$\sigma_{\bar{x}}$
Standard error of the mean
Central limit theorem

Section 7.6
Normal approximation to the binomial distribution
Continuity correction

VIEWPOINT

Why Wait? Apply Now for a College Loan!

The cost of education is high. The cost of not having an education is higher! What kinds of costs can you expect? What about tuition and student fees? What about room and board? What is the total cost for 1 year at college? Perhaps some averages based on random samples of colleges would be useful. For more information, visit the Brase/Brase statistics site at **http://math.college.hmco.com/students** and find the link to the U.S. News site. Then select Education. Search for the geographic regions of the colleges of interest.

CHAPTER REVIEW PROBLEMS

1. Given that x is a normal variable with mean $\mu = 47$ and standard deviation $\sigma = 6.2$, find
   (a) $P(x \leq 60)$ (b) $P(x \geq 50)$ (c) $P(50 \leq x \leq 60)$

2. Given that x is a normal variable with mean $\mu = 110$ and standard deviation $\sigma = 12$, find
   (a) $P(x \leq 120)$ (b) $P(x \geq 80)$ (c) $P(108 \leq x \leq 117)$

3. Find z such that 5% of the area under the standard normal curve lies to the right of z.

4. Find z such that 99% of the area under the standard normal curve lies between $-z$ and z.

5. *Nursing: Exams* On a practical nursing licensing exam, the mean score is 79 and the standard deviation is 9 points.
   (a) What is the standardized score of a student with a raw score of 87?
   (b) What is the standardized score of a student with a raw score of 79?
   (c) Assuming the scores follow a normal distribution, what is the probability that a score selected at random is above 85?

6. *Aptitude Tests: Mechanical* On an auto mechanic aptitude test, the mean score is 270 points and the standard deviation is 35 points.
   (a) If a student has a standardized score of 1.9, how many points is that?
   (b) If a student has a standardized score of -0.25, how many points is that?
   (c) Assuming the scores follow a normal distribution, what is the probability that a student will get between 200 and 340 points?

7. *Recycling: Aluminum Cans* One environmental group did a study of recycling habits in a California community. It found that 70% of the aluminum cans sold in the area were recycled.
   (a) If 400 cans are sold today, what is the probability that 300 or more will be recycled?
   (b) Of the 400 cans sold, what is the probability that between 260 and 300 will be recycled?

8. *Guarantee: Disc Players* Future Electronics makes compact disc players. Its research department found that the life of the laser beam device is normally distributed, with mean 5000 hours and standard deviation 450 hours.
   (a) Find the probability that the laser beam device will wear out in 5000 hours or less.
   (b) *Inverse Normal Distribution* Future Electronics wants to place a guarantee on the players so that no more than 5% fail during the guarantee period. Because the laser pickup is the part most likely to wear out first, the guarantee period will be based on the life of the laser beam device. How many playing hours should the guarantee cover? (Round to the next playing hour.)

9. *Guarantee: Package Delivery* Express Courier Service has found that the delivery time for packages is normally distributed, with mean 14 hours and standard deviation 2 hours.
   (a) For a package selected at random, what is the probability that it will be delivered in 18 hours or less?

(b) *Inverse Normal Distribution* What should be the guaranteed delivery time on all packages for Express Courier to be 95% sure that the package will be delivered before this time? (*Hint:* Note that 5% of the packages will be delivered at a time beyond the guaranteed time period.)

10. *Medical: Blood Type* Blood type AB is found in only 3% of the population (*Textbook of Medical Physiology*, by A. Guyton, M.D.). If 250 people are chosen at random, what is the probability that
    (a) 5 or more will have this blood type?
    (b) between 5 and 10 will have this blood type?

11. *General Discussion* Let x be a random variable representing the amount of sleep each adult in New York City got last night. Consider a sampling distribution of sample means $\bar{x}$.
    (a) As the sample size becomes increasingly large, what distribution does the $\bar{x}$ distribution approach?
    (b) As the sample size becomes increasingly large, what value will the mean $\mu_{\bar{x}}$ of the $\bar{x}$ distribution approach?
    (c) What value will the standard deviation $\sigma_{\bar{x}}$ of the sampling distribution approach?
    (d) How do the two $\bar{x}$ distributions for sample sizes $n = 50$ and $n = 100$ compare?

12. *Drugs: Effects* A new muscle relaxant is available. Researchers from the firm developing the relaxant have done studies that indicate that the time lapse between administration of the drug and beginning effects of the drug is normally distributed, with mean $\mu = 38$ minutes and standard deviation $\sigma = 5$ minutes.
    (a) The drug is administered to one patient selected at random. What is the probability that the time it takes for the drug to go into effect is 35 minutes or less?
    (b) The drug is administered to a random sample of 10 patients. What is the probability that the average time before the drug is effective for all 10 patients is 35 minutes or less?
    (c) Comment on the differences between the results in parts (a) and (b).

13. *Psychology: IQ Scores* Assume that IQ scores are normally distributed with a standard deviation of 15 points and a mean of 100 points. If 100 people are chosen at random, what is the probability that the sample mean of IQ scores will not differ from the population mean by more than 2 points?

14. *Meteorology: Miami and Fairbanks* Let x be a random variable that represents daily high temperatures (in degrees Fahrenheit) in January. The following information is based on a report from the U.S. Department of Commerce Environmental Data Services. For Miami, Florida, the mean of the x distribution is $\mu = 76$, and the standard deviation is approximately $\sigma = 1.9$. For Fairbanks, Alaska, the mean of the x distribution is $\mu = 0$ with approximate standard deviation $\sigma = 5.3$. Assume that x has a normal distribution.
    (a) For one day chosen at random in January, what is the probability that the high temperature in Miami will be less than 77°F? What is the probability that the high temperature in Fairbanks will be less than 3°F?
    (b) If we choose $n = 7$ days in January, what can we say about the probability distribution of $\bar{x}$, the average high temperature? What is the probability that $\bar{x}$ is less than 77°F for Miami? less than 3°F for Fairbanks?

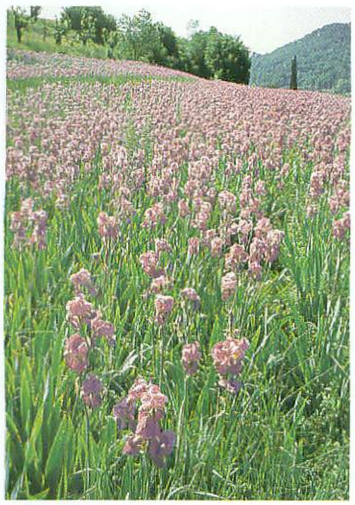
(c) Suppose we cannot assume that x has a normal distribution, but we can say that the distribution is approximately symmetrical and mound-shaped. In this case, what can we say about the $\bar{x}$ probability distribution? If we use all 31 days in January, what is the probability that $\bar{x} < 77°F$ in Miami? less than 3°F in Fairbanks?

DATA HIGHLIGHTS: GROUP PROJECTS

Break into small groups and discuss the following topics. Organize a brief outline in which you summarize the main points of your group discussion.

Iris setosa is a beautiful wildflower that is found in such diverse places as Alaska, the Gulf of St. Lawrence, much of North America, and even in English meadows and parks. R. A. Fisher, with his colleague Dr. Edgar Anderson, studied these flowers extensively. Dr. Anderson described how he collected information on irises:

> I have studied such irises as I could get to see, in as great detail as possible, measuring iris standard after iris standard and iris fall after iris fall, sitting squat-legged with record book and ruler in mountain meadows, in cypress swamps, on lake beaches, and in English parks. [Anderson, E., "The Irises of the Gaspé Peninsula," *Bulletin, American Iris Society,* 59:2–5, 1935.]


Wild iris

The data in Table 7-10 were collected by Dr. Anderson and were published by his friend and colleague R. A. Fisher in a paper entitled "The Use of Multiple Measurements in Taxonomic Problems" (*Annals of Eugenics,* part II, 179–188, 1936). To find these data, visit the Brase/Brase statistics site at **http://math.college.hmco.com/students** and find the link to DASL, the Carnegie Mellon University Data and Story Library. From the DASL site, look under famous data sets.

Let x be a random variable representing petal length. Using a TI-84Plus/TI-83Plus calculator, it was found that the sample mean for the petal length data is $\bar{x} = 1.46$ cm and the sample standard deviation is $s = 0.17$ cm. Figure 7-36 shows a histogram for the given data generated on a TI-84Plus/TI-83Plus calculator.

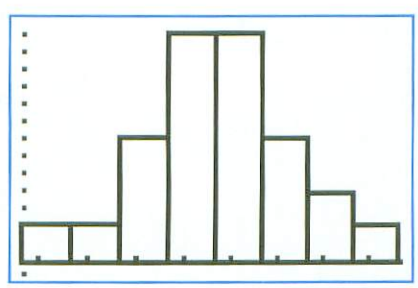
(a) Examine the histogram for petal lengths. Would you say that the distribution is approximately mound-shaped and symmetrical? Our sample has only 50 irises; if many thousands of irises had been used, do you think that the distribution would

TABLE 7-10 Petal Length in Centimeters for *Iris setosa*

1.4	1.4	1.3	1.5	1.4
1.7	1.4	1.5	1.4	1.5
1.5	1.6	1.4	1.1	1.2
1.5	1.3	1.4	1.7	1.5
1.7	1.5	1	1.7	1.9
1.6	1.6	1.5	1.4	1.6
1.6	1.5	1.5	1.4	1.5
1.2	1.3	1.4	1.3	1.5
1.3	1.3	1.3	1.6	1.9
1.4	1.6	1.4	1.5	1.4

FIGURE 7-36

Petal Length (cm) for *Iris setosa* (TI-84Plus/TI-83Plus)



look even more like a normal curve? Let x be the petal length of *Iris setosa*. Research has shown that x has an approximately normal distribution, with mean $\mu = 1.5$ cm and standard deviation $\sigma = 0.2$ cm.

(b) Use the empirical rule with $\mu = 1.5$ and $\sigma = 0.2$ to get an interval in which approximately 68% of the petal lengths will fall. Repeat this for 95% and 99.7%. Examine the raw data and compute the percentage of the raw data that actually falls into each of these intervals (the 68% interval, the 95% interval, and the 99.7% interval). Compare your computed percentages with those given by the empirical rule.

(c) Compute the probability that a petal length is between 1.3 and 1.6 cm. Compute the probability that a petal length is greater than 1.6 cm.

(d) Suppose that a random sample of 30 irises is obtained. Compute the probability that the average petal length for this sample is between 1.3 and 1.6 cm. Compute the probability that the average petal length is greater than 1.6 cm.

(e) Compare your answers for parts (c) and (d). Do you notice any differences? Why would these differences occur?

LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

1. Why are standard z values so important? Is it true that z values have no units of measurement? Why would this be desirable for comparing data sets with *different* units of measurement? How can we assess differences in quality or performance by simply comparing z values under a standard normal curve? Examine the formula to compute standard z values. Notice that it involves *both* the mean and standard deviation. Recall that in Chapter 2 we commented that the mean of a data collection is not entirely adequate to describe the data; you need the standard deviation as well. Discuss this topic again in the light of what you now know about normal distributions and standard z values.

2. Most people would agree that increased information should give better predictions. Discuss how sampling distributions actually enable better predictions by providing more information. Examine Theorem 7.1 again. Suppose that x is a random variable with a *normal* distribution. Then $\bar{x}$, the sample mean based on random samples of size n, also will have a normal distribution for *any* value of $n = 1, 2, 3, \ldots$.

   What happens to the standard deviation of the $\bar{x}$ distribution as n (the sample size) increases? Consider the following table for different values of n.

n	1	2	3	4	10	50	100
$\sigma/\sqrt{n}$	1σ	0.71σ	0.58σ	0.50σ	0.32σ	0.14σ	0.10σ

In this case, "increased information" means a larger sample size n. Give a brief explanation as to why a *large* standard deviation will usually result in poor statistical predictions, whereas a *small* standard deviation usually results in much better predictions. Since the standard deviation of the sampling distribution $\bar{x}$ is $\sigma/\sqrt{n}$, we can decrease the standard deviation by increasing n. In fact, if we look at the preceding table, we see that if we use a sample size of only $n = 4$, we cut the standard deviation of $\bar{x}$ to 50% of the standard deviation σ of x. If we were to use a sample of size $n = 100$, we would cut the standard deviation of $\bar{x}$ to 10% of the standard deviation σ of x.

Give the preceding discussion some thought and explain why you should get much better predictions for μ by using $\bar{x}$ from a sample of size n rather than by just using x. Write a brief essay in which you explain why sampling distributions are an important tool in statistics.

Using Technology

TI-84PLUS/TI-83PLUS • *EXCEL* • *MINITAB* • *SPSS*

As we have seen in this chapter, the value of a sample statistic such as $\bar{x}$ varies from one sample to another. The central limit theorem describes the distribution of the sample statistic $\bar{x}$ when samples are sufficiently large.

We can use technology tools to generate samples of the same size from the same population. Then we can look at the statistic $\bar{x}$ for each sample, and the resulting $\bar{x}$ distribution.

Project Illustrating the Central Limit Theorem

Step 1: Generate random samples of specified size n from a population.

The random-number table enables us to sample from the uniform distribution of digits 0 through 9. Use either the random-number table or a random-number generator to generate 30 samples of size 10.

Step 2: Compute the sample mean $\bar{x}$ of the digits in each sample.

Step 3: Compute the sample mean of the means (i.e., $\bar{x}_{\bar{x}}$) as well as the standard deviation $s_{\bar{x}}$ of the sample means.

The population mean of the uniform distribution of digits from 0 through 9 is 4.5. How does $\bar{x}_{\bar{x}}$ compare to this value?

Step 4: Compare the sample distribution of $\bar{x}$ values to a normal distribution having the mean and standard deviation computed in Step 3.

(a) Use the values of $\bar{x}_{\bar{x}}$ and $s_{\bar{x}}$ computed in Step 3 to create the intervals shown in column 1 of Table 7-11.

(b) Tally the sample means computed in Step 2 to determine how many fall into each interval of column 2. Then compute the percent of data in each interval and record the results in column 3.

(c) The percentages listed in column 4 are those from a normal distribution (see Figure 7-3 showing the empirical rule). Compare the percentages in column 3 to those in column 4. How do the sample percentages compare with the hypothetical normal distribution?

Step 5: Create a histogram showing the sample means computed in Step 2.

TABLE 7-11 Frequency Table of Sample Means

1. Interval	2. Frequency	3. Percent	4. Hypothetical Normal Distribution
$\bar{x} - 3s$ to $\bar{x} - 2s$	Tally the sample means computed in step 2 and place here.	Compute percents from column 2 and place here.	2 or 3%
$\bar{x} - 2s$ to $\bar{x} - s$			13 or 14%
$\bar{x} - s$ to $\bar{x}$			About 34%
$\bar{x}$ to $\bar{x} + s$			About 34%
$\bar{x} + s$ to $\bar{x} + 2s$			13 or 14%
$\bar{x} + 2s$ to $\bar{x} + 3s$			2 or 3%

Look at the histogram, and compare it to a normal distribution with the mean and standard deviation of the $\bar{x}$'s (as computed in Step 3).

Step 6: Compare the results of this project to the central limit theorem.

Increase the sample size of Step 1 to 20, 30, and 40 and repeat Steps 1 to 5.

Technology Hints

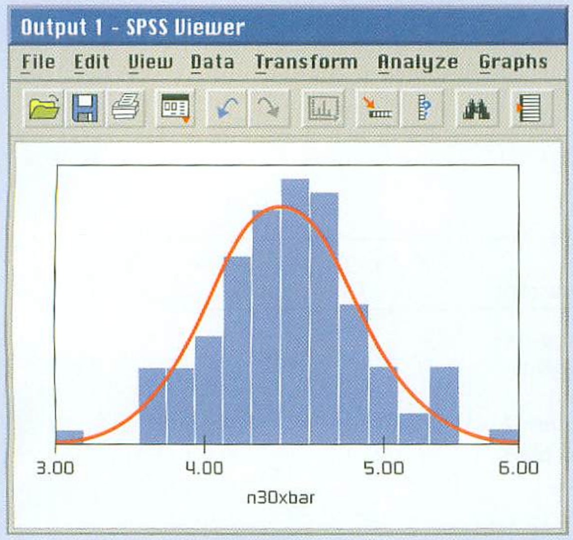
The TI-84Plus and TI-83Plus calculators, Excel, Minitab, and SPSS all support the process of drawing random samples from a variety of distributions. Macros can be written in Excel, Minitab, and the professional version of SPSS to repeat the six steps of the project. Figure 7-37 shows histograms generated by SPSS for random samples of size 30 and size 100. The samples are taken from a uniform probability distribution.

TI-84Plus/TI-83Plus

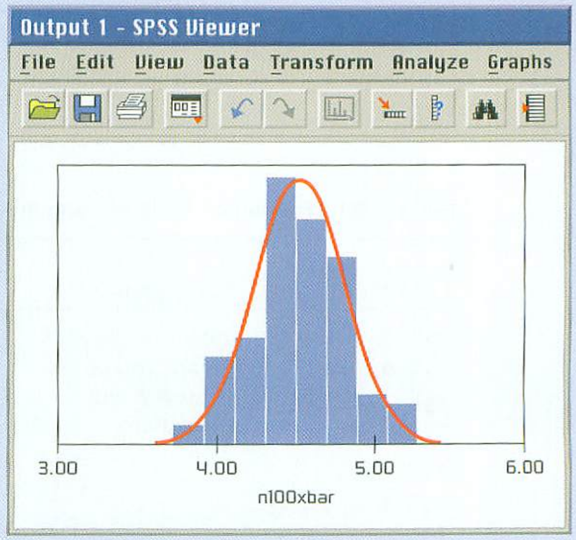
You can generate random samples from uniform, normal, and binomial distributions. Press **MATH** and select **PRB**. Selection **5:randInt(lower, upper, sample size m)** generates m random integers from the specified interval. Selection **6:randNorm(**μ**,** σ**, sample size m)** generates m random numbers from a normal distribution with mean μ and standard deviation σ. Selection **7:randBin(number of trials n, p, sample size m)** generates m random values (number of successes out of n trials) for a binomial distribution with probability of success p on each trial. You can put these values in lists by using **Edit**

FIGURE 7-37 SPSS-Generated Histograms for Samples of Size 30 and Size 100

(a) $n = 30$



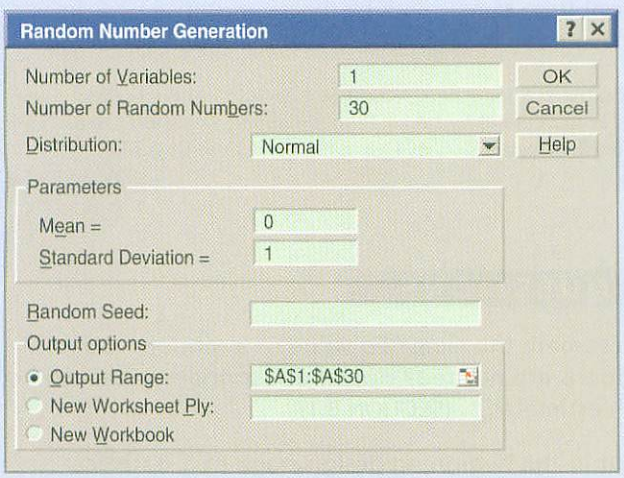
(b) $n = 100$



under **Stat.** Highlight the list header, press Enter, and then select one of the options discussed.

Excel

Use the menu selection **Tools ➤ Data Analysis ➤ Random Number Generator.** The dialogue box provides choices for the population distribution, including uniform, binomial, and normal distributions. Fill in the required parameters and designate the location for the output.



Minitab

Use the menu selection **Calc ➤ Random Data.** Then select the population distribution. The choices include uniform, binomial, and normal distributions. Fill in the dialogue box, where the number of rows indicates the number of data in the sample.

SPSS

SPSS supports random samples from a variety of distributions, including binomial, normal, and uniform. In data view, generate a column of consecutive integers from 1 to n, where n is the sample size. In variable view, name the variables sample1, sample2, and so on through sample30. These variables head the columns containing each of the 30 samples of size n. Then use the menu choices **Transform ➤ Compute.** In the dialogue box, use sample 1 as the target variable for the first sample, and so forth. In the function box, select **RV.UNIFORM(min,max)** for samples from a uniform distribution. Functions **RV.NORMAL(mean,stddev)** and **RV.BINOM(n,p)** provide random samples from normal and binomial distributions, respectively.