CHAPTER 7

The Normal Probability Distribution

Outline

DECISIONS

You are interested in starting your own MENSA-type club. To qualify for the club, the potential member must have intelligence that is in the top 20% of all people. You must decide the baseline score that allows an individual to qualify. See the Decisions project on page 359.

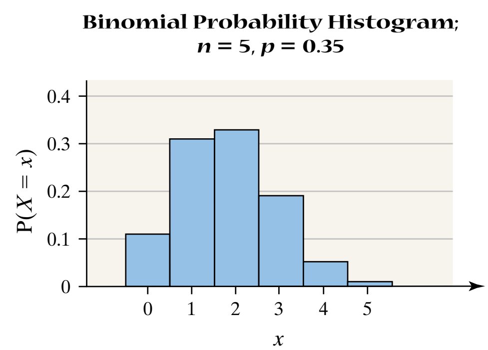
Putting It All Together

In Chapter 6, we introduced discrete probability distributions and, in particular, the binomial probability distribution. We computed probabilities for this discrete distribution using its probability distribution function.

However, we could also determine the probability of any discrete random variable from its probability histogram. For example, the figure shows the probability histogram for the binomial random variable X with $n = 5$ and $p = 0.35$.

From the probability histogram, we can see $P(1) \approx 0.31$. Notice that the width of each rectangle in the probability histogram is 1. Since the area of a rectangle equals height times width, we can think of $P(1)$ as the area of the rectangle corresponding to $X = 1$. Thinking of probability in this fashion makes the transition from computing discrete probabilities to continuous probabilities much easier.

In this chapter, we discuss two continuous distributions, *the uniform distribution* and *the normal distribution*. The greater part of the discussion will focus on the normal distribution, which has many uses and applications.



7.1 Properties of the Normal Distribution

Preparing for This Section Before getting started, review the following:

- Continuous variable (Section 1.1, p. 7)
- Rules for a discrete probability distribution (Section 6.1, p. 287)
- *Z*-score (Section 3.4, pp. 149–150)
- The Empirical Rule (Section 3.2, pp. 131–132)

Objectives

1. **Understand the uniform probability distribution**
2. **Graph a normal curve**
3. **State the properties of the normal curve**
4. **Understand the role of area in the normal density function**
5. **Understand the relation between a normal random variable and a standard normal random variable**

1 Understand the Uniform Probability Distribution

We illustrate a uniform distribution using an example. Using the uniform distribution makes it easy to see the relation between area and probability.

EXAMPLE 1 **Illustrating the Uniform Distribution**

Imagine that a friend of yours is always late. Let the random variable X represent the time from when you are supposed to meet your friend until he shows up. Further suppose that your friend could be on time ($x = 0$) or up to 30 minutes late ($x = 30$), with all 1-minute intervals of times between $x = 0$ and $x = 30$ equally likely. That is, your friend is just as likely to be from 3 to 4 minutes late as he is to be 25 to 26 minutes late. The random variable X can be any value in the interval from 0 to 30, that is, $0 \leq x \leq 30$. Because any two intervals of equal length between 0 and 30, inclusive, are equally likely, the random variable X is said to follow a **uniform probability distribution**.

When we compute probabilities for discrete random variables, we usually substitute the value of the random variable into a formula.

Things are not as easy for continuous random variables. Since there are an infinite number of possible outcomes for continuous random variables the probability of observing a particular value of a continuous random variable is zero. For example, the probability that your friend is exactly 12.9438823 minutes late is zero. This result is based on the fact that classical probability is found by dividing the number of ways an event can occur by the total number of possibilities. There is one way to observe 12.9438823, and there are an infinite number of possible values between 0 and 30, so we get a probability that is zero. To resolve this problem, we compute probabilities of continuous random variables over an interval of values. For example, we might compute the probability that your friend is between 10 and 15 minutes late. To find probabilities for continuous random variables, we use *probability density functions*.

Definition

A **probability density function** is an equation used to compute probabilities of continuous random variables that must satisfy the following two properties.

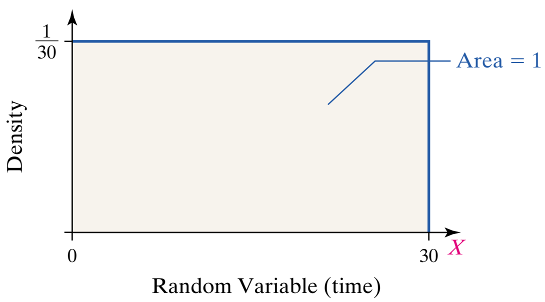
1. The total area under the graph of the equation over all possible values of the random variable must equal 1.
2. The height of the graph of the equation must be greater than or equal to 0 for all possible values of the random variable. That is, the graph of the equation must lie on or above the horizontal axis for all possible values of the random variable.

In Other Words
To find probabilities for continuous random variables, we do not use probability distribution functions (as we did for discrete random variables). Instead, we use probability density functions. The word *density* is used because it refers to the number of individuals per unit of area.

Property 1 is similar to the rule for discrete probability distributions that stated the sum of the probabilities must add up to 1. Property 2 is similar to the rule that stated all probabilities must be greater than or equal to 0.

Figure 1 illustrates the properties for the example about your friend who is always late. Since all possible values of the random variable between 0 and 30 are equally likely, the graph of the probability density function for uniform random variables is a rectangle. Because the random variable is any number between 0 and 30 inclusive, the width of the rectangle is 30. Since the area under the graph of the probability density function must equal 1, and the area of a rectangle equals height times width, the height of the rectangle must be $\frac{1}{30}$.

Figure 1
Uniform Density Function



A pressing question remains: How do we use density functions to find probabilities of continuous random variables?

The area under the graph of a density function over some interval represents the probability of observing a value of the random variable in that interval.

The following example illustrates this statement.

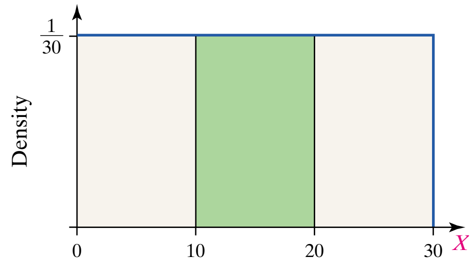
EXAMPLE 2 Area as a Probability

Problem: Refer to the situation presented in Example 1. What is the probability that your friend will be between 10 and 20 minutes late the next time you meet him?

Approach: Figure 1 presented the graph of the density function. We need to find the area under the graph between 10 and 20 minutes.

Solution: Figure 2 presents the graph of the density function with the area we wish to find shaded in green.

Figure 2



The width of the rectangle is 10 and its height is $\frac{1}{30}$. Therefore, the area between 10 and 20 is $10\left(\frac{1}{30}\right) = \frac{1}{3}$. The probability that your friend is between 10 and 20 minutes late is $\frac{1}{3}$.

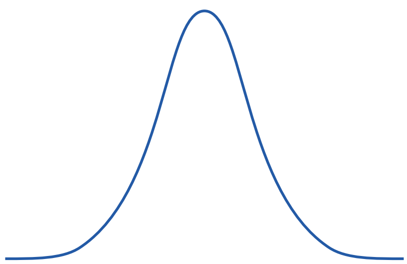
Now Work Problem 13.

We introduced the uniform density function so that we could associate probability with area. We are now better prepared to discuss the most popular continuous distribution, the normal distribution.

2 Graph a Normal Curve

Many continuous random variables, such as IQ scores, birth weights of babies, or weights of M&Ms, have relative frequency histograms that have a shape similar to Figure 3. Relative frequency histograms that have a shape similar to Figure 3 are said to have the shape of a **normal curve**.

Figure 3



Definition

A continuous random variable is **normally distributed** or has a **normal probability distribution** if its relative frequency histogram of the random variable has the shape of a normal curve.

Figure 4

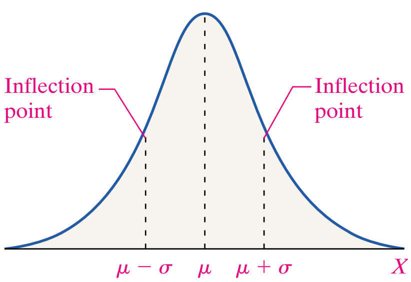


Figure 4 shows a normal curve, demonstrating the role μ and σ play in drawing the curve. Look back at Figure 5 on page 113 in Section 3.1. For any distribution, the mode represents the "high point" of the graph of the distribution. The median represents the point where 50% of the area under the distribution is to the left and 50% of the area under the distribution is to the right. The mean represents the balancing point of the graph of the distribution (see Figure 2 on page 109 in Section 3.1). For symmetric distributions, such as the normal distribution, the mean = median = mode. Because of this, the mean, μ, is the "high point" of the graph of the distribution.

The points at $x = \mu - \sigma$ and $x = \mu + \sigma$ are the *inflection points* on the normal curve. The **inflection points** are the points on the curve where the curvature

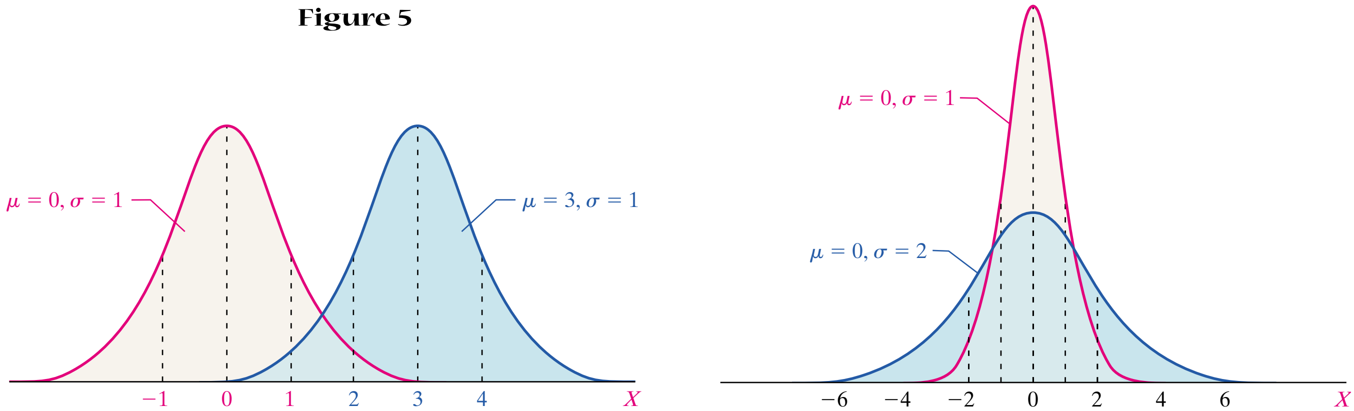
Historical Note

Karl Pearson coined the phrase *normal curve*. He did not do this to imply that a distribution that is not normal is *abnormal*. Rather, Pearson wanted to avoid giving the name of the distribution a proper name, such as Gaussian (as in Carl Friedrich Gauss).

of the graph changes. To the left of $x = \mu - \sigma$ and to the right of $x = \mu + \sigma$, the curve is drawn upward $(\smile)$ or $(\smile)$. In between $x = \mu - \sigma$ and $x = \mu + \sigma$, the curve is drawn downward $(\frown)$.*

Figure 5 shows how changes in μ and σ change the position or shape of a normal curve. In Figure 5(a), two normal density curves are drawn with the location of the inflection points labeled. One density curve has $\mu = 0, \sigma = 1$, and the other has $\mu = 3, \sigma = 1$. We can see that increasing the mean from 0 to 3 caused the graph to shift three units to the right. In Figure 5(b), two normal density curves are drawn, again with the inflection points labeled. One density curve has $\mu = 0, \sigma = 1$, and the other has $\mu = 0, \sigma = 2$. We can see that increasing the standard deviation from 1 to 2 causes the graph to become flatter and more spread out.

Figure 5



Now Work Problem 25.

3 State the Properties of the Normal Curve

The normal probability density function satisfies all the requirements that are necessary to have a probability distribution. We list the properties of the normal density curve next.

Historical Note

Abraham de Moivre was born in France on May 26, 1667. He is known as a great contributor to the areas of probability and trigonometry. In 1685, he moved to England. De Moivre was elected a fellow of the Royal Society in 1697. He was part of the commission to settle the dispute between Newton and Leibniz regarding who was the discoverer of calculus. He published *The Doctrine of Chance* in 1718. In 1733, he developed the equation that describes the normal curve. Unfortunately, de Moivre had a difficult time being accepted in English society (perhaps due to his accent) and was able to make only a meager living tutoring mathematics. An interesting piece of information regarding de Moivre; he correctly predicted the day of his death, November 27, 1754.

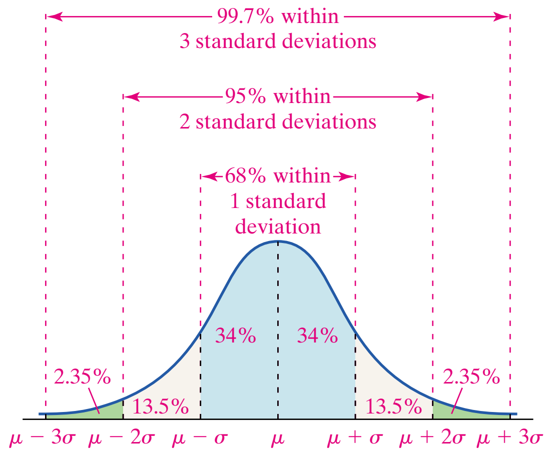
Properties of the Normal Density Curve

1. It is symmetric about its mean, μ.
2. Because mean = median = mode, the highest point occurs at $x = \mu$.
3. It has inflection points at $\mu - \sigma$ and $\mu + \sigma$.
4. The area under the curve is 1.
5. The area under the curve to the right of μ equals the area under the curve to the left of μ, which equals $\frac{1}{2}$.
6. As x increases, without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As x decreases without bound (gets larger and larger in the negative direction), the graph approaches, but never reaches, the horizontal axis.
7. The Empirical Rule: Approximately 68% of the area under the normal curve is between $x = \mu - \sigma$ and $x = \mu + \sigma$. Approximately 95% of the area under the normal curve is between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$. Approximately 99.7% of the area under the normal curve is between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$. See Figure 6.

*The vertical scale on the graph, which indicates density, is purposely omitted. The vertical scale, while important, will not play a role in any of the computations using this curve.

Figure 6

Normal Distribution



4 Understand the Role of Area in the Normal Density Function

Let's look at an example of a normally distributed random variable.

EXAMPLE 3 A Normal Random Variable

Problem: The relative frequency distribution given in Table 1 represents the heights of a pediatrician's 200 three-year-old female patients. The raw data indicate that the mean height of the patients is $\mu = 38.72$ inches with standard deviation $\sigma = 3.17$ inches.

(a) Draw a relative frequency histogram of the data. Comment on the shape of the distribution.

(b) Draw a normal curve with $\mu = 38.72$ inches and $\sigma = 3.17$ inches on the relative frequency histogram. Compare the area of the rectangle for heights between 40 and 40.9 inches to the area under the normal curve for heights between 40 and 40.9 inches.

Table 1

Height (Inches)	Relative Frequency
29.0–29.9	0.005
30.0–30.9	0.005
31.0–31.9	0.005
32.0–32.9	0.025
33.0–33.9	0.02
34.0–34.9	0.055
35.0–35.9	0.075
36.0–36.9	0.09
37.0–37.9	0.115
38.0–38.9	0.15
39.0–39.9	0.12
40.0–40.9	0.11
41.0–41.9	0.07
42.0–42.9	0.06
43.0–43.9	0.035
44.0–44.9	0.025
45.0–45.9	0.025
46.0–46.9	0.005
47.0–47.9	0.005

Approach

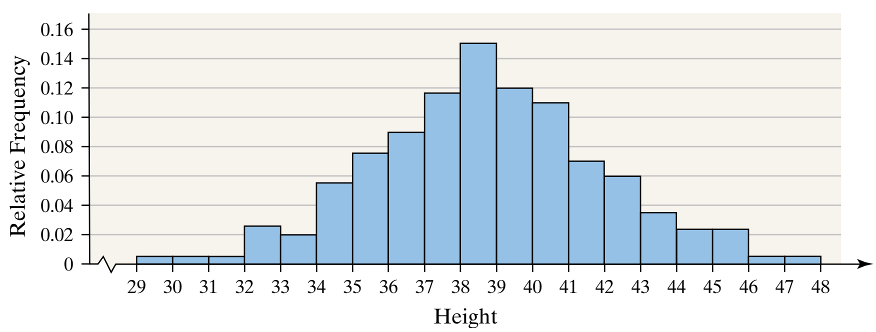
(a) Draw the relative frequency histogram. If the histogram looks like Figure 4, we say that height is approximately normal. We say "approximately normal," rather than "normal," because the normal curve is an "idealized" description of the data and data rarely follows the curve exactly.

(b) Draw the normal curve on the histogram with the high point at μ and the inflection points at $\mu - \sigma$ and $\mu + \sigma$. Shade the rectangle corresponding to heights between 40 and 40.9 inches, and compare the area of the shaded region to the area under the normal curve between 40 and 40.9.

Solution

(a) Figure 7 shows the relative frequency distribution. The relative frequency histogram is symmetric and bell shaped.

Figure 7 Height of Three-Year-Old Females

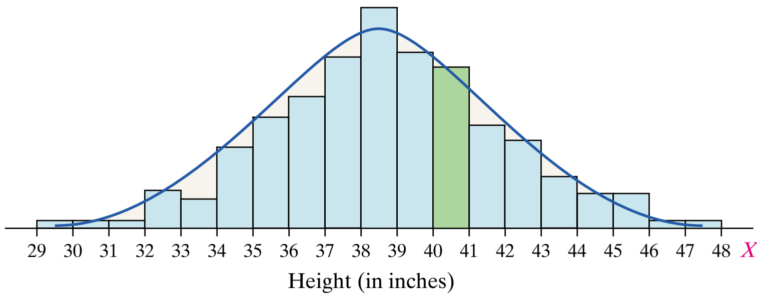


CAUTION
It is rare for a continuous random variable to be exactly normal. Therefore, we usually say that a random variable is approximately normal if its histogram is bell shaped and symmetric.

(b) The normal curve with $\mu = 38.72$ and $\sigma = 3.17$ is superimposed on the relative frequency histogram in Figure 8. The figure demonstrates that the normal curve describes the heights of 3-year-old girls fairly well. We conclude that the heights of 3-year-old girls are approximately normal with $\mu = 38.72$ and $\sigma = 3.17$.

Figure 8 also shows the rectangle corresponding to heights between 40 and 40.9 inches. The area of this rectangle represents the proportion of 3-year-old females between 40 and 40.9 inches. Notice that the area of this shaded region is very close to the area under the normal curve for the same region, so we can use the area under the normal curve to approximate the proportion of 3-year-old females with heights between 40 and 40.9 inches!

Figure 8
Heights of 3-Year-Old Female Patients



In Other Words
Models are not always mathematical. For example, a map can be thought of as a model of a highway system. The model does not show every detail of the highway system (such as traffic lights), but it does serve the purpose of describing how to get from point A to point B. Mathematical models do the same thing: They make assumptions to simplify the mathematics, while still trying to accomplish the goal of accurately describing reality.

The normal curve drawn in Figure 8 is a *model*. In mathematics, a **model** is an equation, table, or graph that is used to describe reality. The normal distribution or normal curve is a model that is used to describe variables that are approximately normally distributed. For example, we saw in Example 3(b) that the normal curve drawn in Figure 8 does a good job of describing the observed distribution of heights of 3-year-old females.

The equation (model) that is used to determine the probability of continuous random variable is called a **probability density function** (or **pdf**). The **normal probability density function** is given by

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

where μ is the mean and σ is the standard deviation of the normal random variable. This equation represents the normal distribution. Don't feel threatened by this equation because we will not be using it in this course. Instead, we will use the normal distribution in graphical form by drawing the normal curve, as we did in Figure 5.

We now summarize the role area plays in the normal curve.

The Area under a Normal Curve

Suppose a random variable X is normally distributed with mean μ and standard deviation σ. The area under the normal curve for any interval of values of the random variable X represents either

- the proportion of the population with the characteristic described by the interval of values or
- the probability that a randomly selected individual from the population will have the characteristic described by the interval of values.

In Other Words
The area under a normal curve is a proportion or probability.

EXAMPLE 4 Interpreting the Area under a Normal Curve

Historical Note

The normal probability distribution is often referred to as the Gaussian distribution in honor of Carl Gauss, the individual thought to have discovered the idea. However, it was actually Abraham de Moivre who first wrote down the equation of the normal distribution. Gauss was born in Brunswick, Germany, on April 30, 1777. Mathematical prowess was evident early in Gauss's life. At age 8 he was able to instantly add the first 100 integers. In 1799, Gauss earned his doctorate. The subject of his dissertation was the Fundamental Theorem of Algebra. In 1809, Gauss published a book on the mathematics of planetary orbits. In this book, he further developed the theory of least-squares regression by analyzing the errors. The analysis of these errors led to the discovery that errors follow a normal distribution. Gauss was considered to be "glacially cold" as a person and had troubled relationships with his family. Gauss died on February 23, 1855.

Problem: The serum total cholesterol for males 20 to 29 years old is approximately normally distributed with mean $\mu = 180$ and $\sigma = 36.2$ based on data obtained from the National Health and Nutrition Examination Survey.

(a) Draw a normal curve with the parameters labeled.

(b) An individual with total cholesterol greater than 200 is considered to have high cholesterol. Shade the region under the normal curve to the right of $X = 200$.

(c) Suppose the area under the normal curve to the right of $X = 200$ is 0.2903. (You will learn how to find this area in Section 7.3.) Provide two interpretations of this result.

Approach: **(a)** Draw the normal curve with the mean $\mu = 180$ labeled at the high point and the inflection points at $\mu - \sigma = 180 - 36.2 = 143.8$ and $\mu + \sigma = 180 + 36.2 = 216.2$.

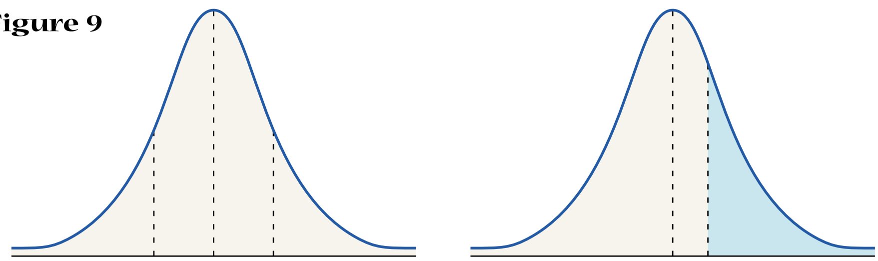
(b) Shade the region under the normal curve to the right of $X = 200$.

(c) The two interpretations of this shaded region are (1) the proportion of 20- to 29-year-old males who have high cholesterol and (2) the probability that a randomly selected 20- to 29-year-old male has high cholesterol.

Solution

(a) Figure 9(a) shows the graph of the normal curve.

Figure 9



(b) Figure 9(b) shows the region under the normal curve to the right of $X = 200$ shaded.

(c) The two interpretations for the area of this shaded region are (1) the proportion of 20- to 29-year-old males that have high cholesterol is 0.2903 and (2) the probability that a randomly selected 20- to 29-year-old male has high cholesterol is 0.2903.

Now Work Problems 29 and 33.

5 Understand the Relation between a Normal Random Variable and a Standard Normal Random Variable

At this point, we know that a random variable X is approximately normally distributed if its relative frequency histogram has the shape of a normal curve. We use a normal random variable with mean μ and standard deviation σ to model the distribution of X. The area below the normal curve (model of X) represents the proportion of the population with a given characteristic or the probability that a randomly selected individual from the population will have a given characteristic.

The question now becomes, "How do I find the area under the normal curve?" Finding the area under a curve requires techniques introduced in calculus, which are beyond the scope of this text. An alternative would be to use a series of tables to find areas. However, this would result in an infinite number of tables being created for each possible mean and standard deviation!

A solution to the problem lies in the Z-score. Recall that the Z-score allows us to transform a random variable X with mean μ and standard deviation σ into a random variable Z with mean 0 and standard deviation 1.

In Other Words
The term *normal* refers to the shape of the distribution of a normal random variable.

Standardizing a Normal Random Variable

Suppose the random variable X is normally distributed with mean μ and standard deviation σ. Then the random variable

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

is normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1$. The random variable Z is said to have the **standard normal distribution**.

This result is powerful! We need only one table of areas corresponding to the standard normal distribution. If a normal random variable has mean different from 0 or standard deviation different from 1, we transform the normal random variable into a standard normal random variable Z, and then we use a table to find the area and, therefore, the probability.

We demonstrate the idea behind standardizing a normal random variable in the next example.

In Other Words
To find the area under any normal curve, we first find the Z-score of the normal random variable. Then we use a table to find the area.

EXAMPLE 5 Relation between a Normal Random Variable and a Standard Normal Random Variable

Problem: The heights of a pediatrician's 200 three-year-old female patients are approximately normal with mean $\mu = 38.72$ inches and $\sigma = 3.17$ inches. We wish to demonstrate that the area under the normal curve between 35 and 38 inches is equal to the area under the standard normal curve between the Z-scores corresponding to heights of 35 and 38 inches.

Approach

Step 1: Draw a normal curve and shade the area representing the proportion of 3-year-old females between 35 and 38 inches tall.

Step 2: Standardize the random variable $X = 35$ and $X = 38$ using

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

Step 3: Draw the standard normal curve with the standardized versions of $X = 35$ and $X = 38$ labeled. Shade the area that represents the proportion of 3-year-old females between 35 and 38 inches tall. Comment on the relation between the two shaded regions.

Solution

Step 1: Figure 10(a) shows the normal curve with mean $\mu = 38.72$ and $\sigma = 3.17$. The region between $X = 35$ and $X = 38$ is shaded.

Step 2: With $\mu = 38.72$ and $\sigma = 3.17$, the standardized version of $X = 35$ is

$$Z = \frac{X - \mu}{\sigma} = \frac{35 - 38.72}{3.17} = -1.17$$

The standardized version of $X = 38$ is

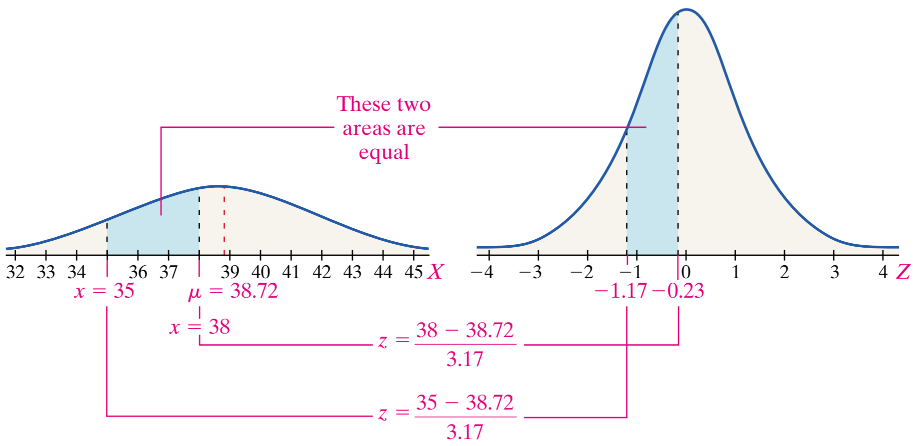
$$Z = \frac{X - \mu}{\sigma} = \frac{38 - 38.72}{3.17} = -0.23$$

CAUTION
Recall that we round Z-scores to two decimal places.

***Step 3*:** Figure 10(b) shows the standard normal curve with the region between $Z = -1.17$ and $Z = -0.23$ shaded.

Figure 10

(a) Normal Curve with $\mu = 38.72$ and $\sigma = 3.17$
(b) Standard Normal Curve



(a) **(b)**

The area under the normal curve with $\mu = 38.72$ inches and $\sigma = 3.17$ inches bounded to the left by $X = 35$ and bounded to the right by $X = 38$ is equal to the area under the standard normal curve bounded to the left by $Z = -1.17$ and bounded to the right by $Z = -0.23$.

Now Work Problem 35.

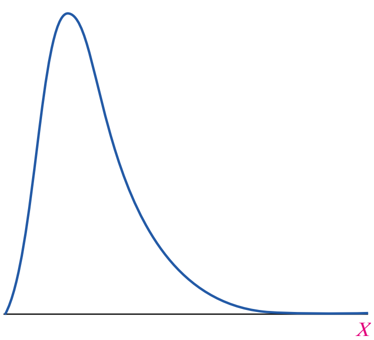
7.1 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

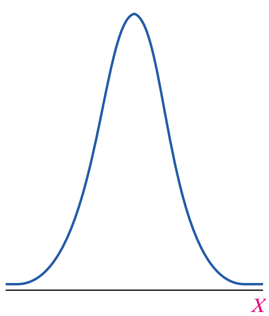
1. State the two characteristics of the graph of a probability density function.
2. To find the probabilities for continuous random variables, we do not use probability _____ functions, but instead we use probability _____ functions.
3. Provide two interpretations of the area under the graph of a probability density function.
4. Why do we standardize normal random variables to find the area under any normal curve?
5. The points at $x =$ _____ and $x =$ _____ are the inflection points on the normal curve.
6. As σ increases, the normal density curve becomes more spread out. Knowing the area under the density curve must be 1, what effect does increasing σ have on the height of the curve?

For Problems 7–12, determine whether the graph can represent a normal density function. If it cannot, explain why.

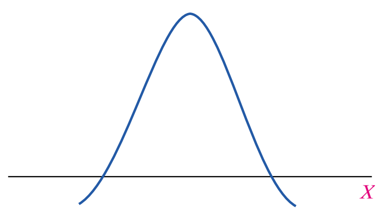
7.



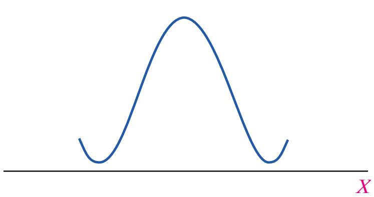
8.



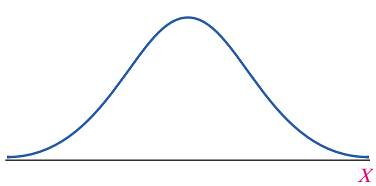
9.



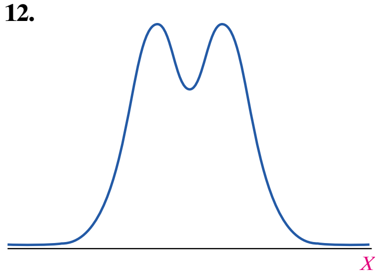
10.



11.



12.



Skill Building

Problems 13–16 use the information presented in Examples 1 and 2.

13. NW Find the probability that your friend is between 5 and 10 minutes late.

14. Find the probability that your friend is between 15 and 25 minutes late.

15. Find the probability that your friend is at least 20 minutes late.

16. Find the probability that your friend is no more than 5 minutes late.

17. **Uniform Distribution** The random-number generator on calculators randomly generates a number between 0 and 1. The random variable X, the number generated, follows a uniform probability distribution.

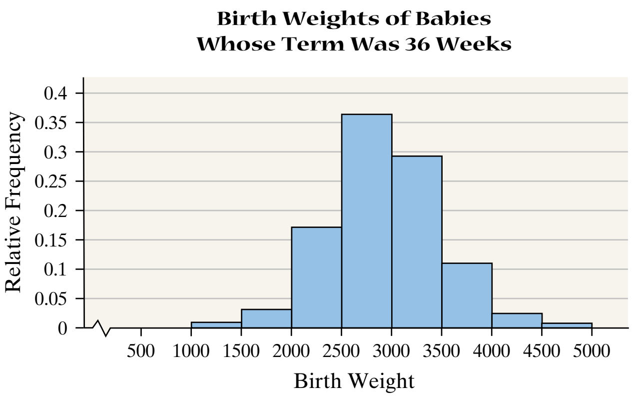
(a) Draw the graph of the uniform density function.
(b) What is the probability of generating a number between 0 and 0.2?
(c) What is the probability of generating a number between 0.25 and 0.6?
(d) What is the probability of generating a number greater than 0.95?
(e) Use your calculator or statistical software to randomly generate 200 numbers between 0 and 1. What proportion of the numbers are between 0 and 0.2? Compare the result with part (b).

18. **Uniform Distribution** Suppose the reaction time X (in minutes) of a certain chemical process follows a uniform probability distribution with $5 \leq X \leq 10$.

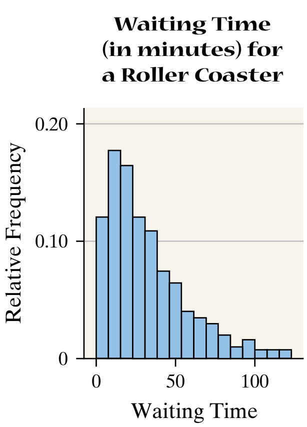
(a) Draw the graph of the density curve.
(b) What is the probability that the reaction time is between 6 and 8 minutes?
(c) What is the probability that the reaction time is between 5 and 8 minutes?
(d) What is the probability that the reaction time is less than 6 minutes?

In Problems 19–22, determine whether or not the histogram indicates that a normal distribution could be used as a model for the variable.

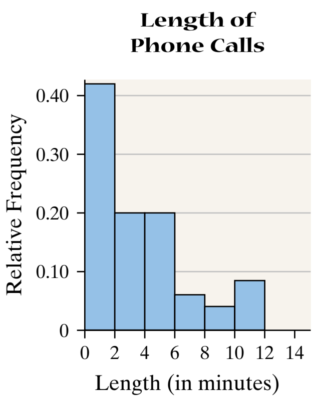
19. **Birth Weights** The following relative frequency histogram represents the birth weights (in grams) of babies whose term was 36 weeks.



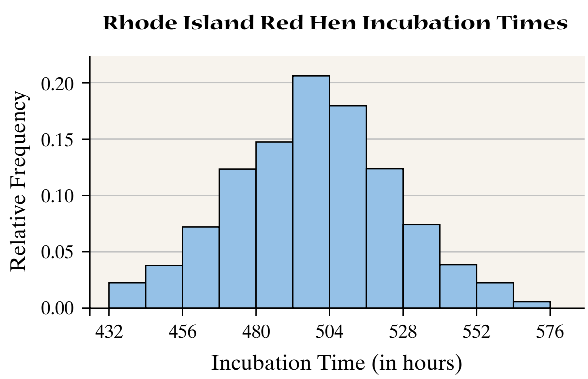
20. **Waiting in Line** The following relative frequency histogram represents the waiting time in line (in minutes) for the Demon Roller Coaster for 2000 randomly selected people on a Saturday afternoon in the summer.



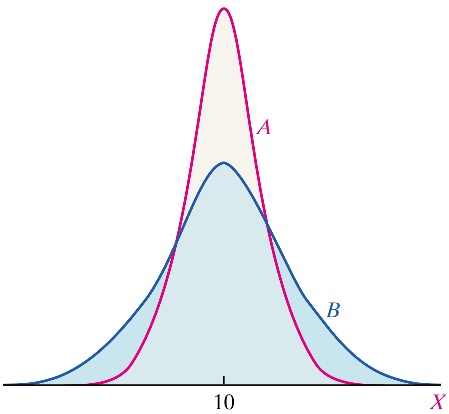
21. **Length of Phone Calls** The following relative frequency histogram represents the length of phone calls on my wife's cell phone during the month of September.



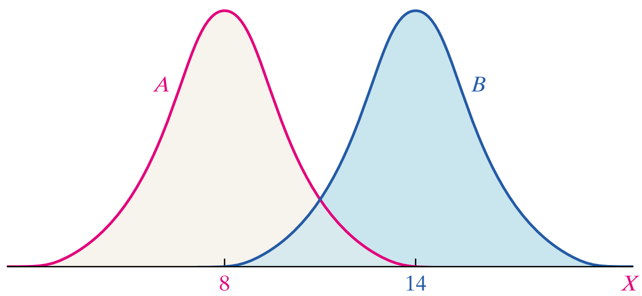
22. **Incubation Times** The following relative frequency histogram represents the incubation times of a random sample of Rhode Island Red Hens' eggs.



23. One graph in the following figure represents a normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 3$. The other graph represents a normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 2$. Determine which graph is which and explain how you know.

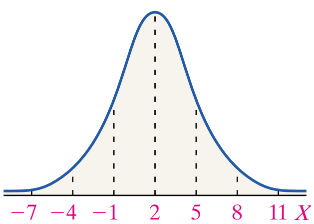


24. One graph in the following figure represents a normal distribution with mean $\mu = 8$ and standard deviation $\sigma = 2$. The other graph represents a normal distribution with mean $\mu = 14$ and standard deviation $\sigma = 2$. Determine which graph is which and explain how you know.

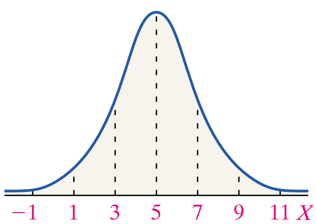


In Problems 25–28, the graph of a normal curve is given. Use the graph to identify the value of μ and σ.

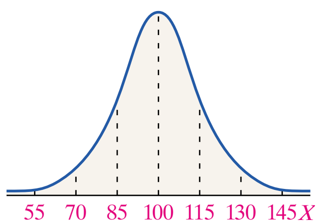
25. NW



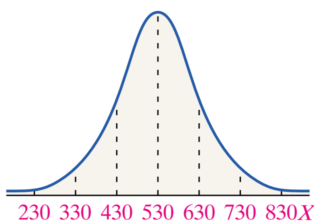
26.



27.



28.



Applying the Concepts

29. Cell Phone Rates NW Suppose the monthly charge for cell phone plans in the United States is normally distributed with mean $\mu = \$62$ and standard deviation $\sigma = \$18$. (*Source*: Based on information obtained from *Consumer Reports*)

(a) Draw a normal curve with the parameters labeled.
(b) Shade the region that represents the proportion of plans that charge less than \$44.
(c) Suppose the area under the normal curve to the left of $X = \$44$ is 0.1587. Provide two interpretations of this result.

30. Refrigerators Suppose the life of refrigerators is normally distributed with mean $\mu = 14$ years and standard deviation $\sigma = 2.5$ years. (*Source*: Based on information obtained from *Consumer Reports*)

(a) Draw a normal curve with the parameters labeled.
(b) Shade the region that represents the proportion of refrigerators that are kept for more than 17 years.
(c) Suppose the area under the normal curve to the right of $X = 17$ is 0.1151. Provide two interpretations of this result.

31. Birth Weights The birth weights of full-term babies are normally distributed with mean $\mu = 3400$ grams and $\sigma = 505$ grams. (*Source*: Based on data obtained from the *National Vital Statistics Report,* Vol. 48, No. 3)

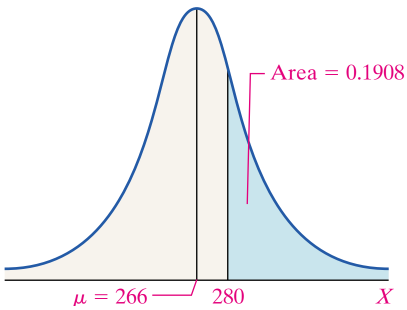
(a) Draw a normal curve with the parameters labeled.
(b) Shade the region that represents the proportion of full-term babies who weigh more than 4410 grams.
(c) Suppose the area under the normal curve to the right of $X = 4410$ is 0.0228. Provide two interpretations of this result.

32. Height of 10-Year-Old Males The heights of 10-year-old males are normally distributed with mean $\mu = 55.9$ inches and $\sigma = 5.7$ inches.

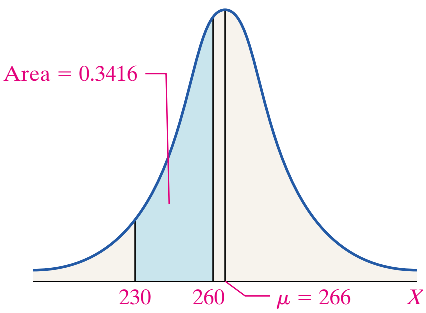
(a) Draw a normal curve with the parameters labeled.
(b) Shade the region that represents the proportion of 10-year-old males who are less than 46.5 inches tall.
(c) Suppose the area under the normal curve to the left of $X = 46.5$ is 0.0496. Provide two interpretations of this result.

33. Gestation Period NW The lengths of human pregnancy are normally distributed with $\mu = 266$ days and $\sigma = 16$ days.

(a) The following figure represents the normal curve with $\mu = 266$ days and $\sigma = 16$ days. The area to the right of $X = 280$ is 0.1908. Provide two interpretations of this area.

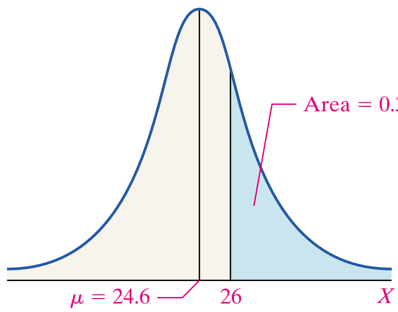


(b) The following figure represents the normal curve with $\mu = 266$ days and $\sigma = 16$ days. The area between $X = 230$ and $X = 260$ is 0.3416. Provide two interpretations of this area.

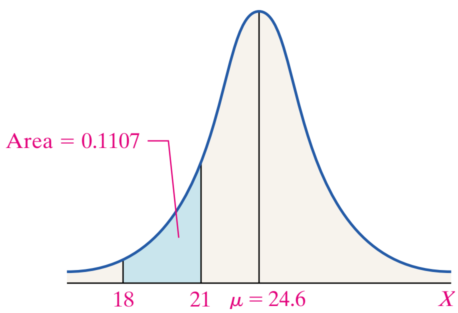


34. Miles per Gallon Elena conducts an experiment in which she fills up the gas tank on her Toyota Camry 40 times and records the miles per gallon for each fill-up. A histogram of the miles per gallon indicates that a normal distribution with mean of 24.6 miles per gallon and a standard deviation of 3.2 miles per gallon could be used to model the gas mileage for her car.

(a) The following figure represents the normal curve with $\mu = 24.6$ miles per gallon and $\sigma = 3.2$ miles per gallon. The area under the curve to the right of $X = 26$ is 0.3309. Provide two interpretations of this area.



(b) The following figure represents the normal curve with $\mu = 24.6$ miles per gallon and $\sigma = 3.2$ miles per gallon. The area under the curve between $X = 18$ and $X = 21$ is 0.1107. Provide two interpretations of this area.



35. NW A random variable X is normally distributed with $\mu = 10$ and $\sigma = 3$.

(a) Compute $Z_1 = \dfrac{X_1 - \mu}{\sigma}$ for $X_1 = 8$.

(b) Compute $Z_2 = \dfrac{X_2 - \mu}{\sigma}$ for $X_2 = 12$.

(c) The area under the normal curve between $X_1 = 8$ and $X_2 = 12$ is 0.495. What is the area between Z_1 and Z_2?

36. A random variable X is normally distributed with $\mu = 25$ and $\sigma = 6$.

(a) Compute $Z_1 = \dfrac{X_1 - \mu}{\sigma}$ for $X_1 = 18$.

(b) Compute $Z_2 = \dfrac{X_2 - \mu}{\sigma}$ for $X_2 = 30$.

(c) The area under the normal curve between $X_1 = 18$ and $X_2 = 30$ is 0.6760. What is the area between Z_1 and Z_2?

37. Hitting a Pitching Wedge In the game of golf, distance control is just as important as how far a player hits the ball. Suppose Michael went to the driving range with his range finder and hit 75 golf balls with his pitching wedge and measured the distance each ball traveled (in yards). He obtained the following data:

100	97	101	101	103	100	99	100	100
104	100	101	98	100	99	99	97	101
104	99	101	101	101	100	96	99	99
98	94	98	107	98	100	98	103	100
98	94	104	104	98	101	99	97	103
102	101	101	100	95	104	99	102	95
99	102	103	97	101	102	96	102	99
96	108	103	100	95	101	103	105	100
94	99	95						

(a) Use MINITAB or some other statistical software to construct a relative frequency histogram. Comment on the shape of the distribution.
(b) Use MINITAB or some other statistical software to draw the normal density function on the relative frequency histogram.
(c) Do you think the normal density function accurately describes the distance Michael hits a pitching wedge? Why?

38. Heights of Five-Year-Old Females The following frequency distribution represents the heights (in inches) of eighty randomly selected five-year-old females.

44.5	42.4	42.2	46.2	45.7	44.8	43.3	39.5
45.4	43.0	43.4	44.7	38.6	41.6	50.2	46.9
39.6	44.7	36.5	42.7	40.6	47.5	48.4	37.5
45.5	43.3	41.2	40.5	44.4	42.6	42.0	40.3
42.0	42.2	38.5	43.6	40.6	45.0	40.7	36.3
44.5	37.6	42.2	40.3	48.5	41.6	41.7	38.9
39.5	43.6	41.3	38.8	41.9	40.3	42.1	41.9
42.3	44.6	40.5	37.4	44.5	40.7	38.2	42.6
44.0	35.9	43.7	48.1	38.7	46.0	43.4	44.6
37.7	34.6	42.4	42.7	47.0	42.8	39.9	42.3

(a) Use MINITAB or some other statistical software to construct a relative frequency histogram. Comment on the shape of the distribution.
(b) Use MINITAB or some other statistical software to draw the normal density function on the relative frequency histogram.
(c) Do you think the normal density function accurately describes the heights of 5-year-old females? Why?

7.2 The Standard Normal Distribution

Preparing for This Section Before getting started, review the following:

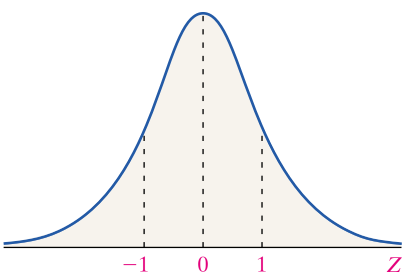
- The Complement Rule (Section 5.2, pp. 244–245)

Objectives

1. **Find the area under the standard normal curve**
2. **Find Z-scores for a given area**
3. **Interpret the area under the standard normal curve as a probability**

In Section 7.1, we introduced the normal distribution. We learned that if X is a normally distributed random variable, we can use the area under the normal density function to obtain the proportion of a population, or the probability that a randomly selected individual from the population has a certain characteristic. To find the area under the normal curve, we first convert the random variable X to a standard normal random variable Z with mean $\mu = 0$ and standard deviation $\sigma = 1$ and find the area under the standard normal curve. This section discusses methods for finding the area under the standard normal curve.

Figure 11



Properties of the Standard Normal Distribution

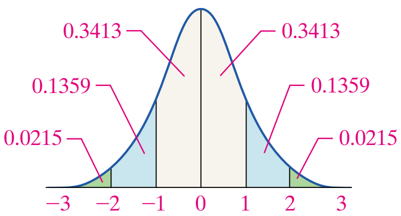
The standard normal distribution has a mean of 0 and a standard deviation of 1. The standard normal curve, therefore, will have its high point located at 0 and inflection points located at −1 and +1. We use the random variable Z to represent a standard normal random variable. The graph of the standard normal curve is presented in Figure 11.

Although we stated the properties of normal curves in Section 7.1, it is worthwhile to restate them here in terms of the standard normal curve.

Properties of the Standard Normal Curve

1. It is symmetric about its mean, $\mu = 0$ and has standard deviation $\sigma = 1$.
2. The mean = median = mode = 0. Its highest point occurs at $\mu = 0$.
3. It has inflection points at $\mu - \sigma = 0 - 1 = -1$ and $\mu + \sigma = 0 + 1 = 1$.
4. The area under the curve is 1.
5. The area under the curve to the right of $\mu = 0$ equals the area under the curve to the left of $\mu = 0$, which equals $\frac{1}{2}$.
6. As Z increases, the graph approaches, but never equals, zero. As Z decreases, the graph approaches, but never equals, zero.
7. The Empirical Rule: Approximately $0.68 = 68\%$ of the area under the standard normal curve is between -1 and 1. Approximately $0.95 = 95\%$ of the area under the standard normal curve is between -2 and 2. Approximately $0.997 = 99.7\%$ of the area under the standard normal curve is between -3 and 3. See Figure 12.

Figure 12



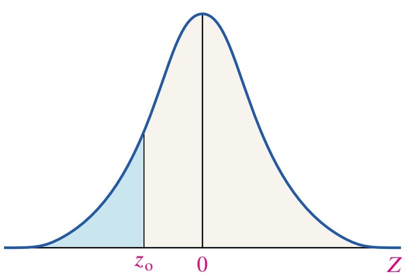
We now discuss the procedure for finding area under the standard normal curve.

1 Find the Area under the Standard Normal Curve

We discuss two methods for finding area under the standard normal curve. The first method uses a table of areas that has been constructed for various values of Z. The second method involves the use of statistical software or a calculator with advanced statistical features.

Table IV, which can be found in the inside back cover of the text or in Appendix A, gives areas under the standard normal curve for values to the left of a specified Z-score, z_0, as shown in Figure 13.

Figure 13



The shaded region represents the area under the standard normal curve to the left of $Z = z_0$. Whenever finding area under a normal curve, you should sketch a normal curve and shade the area you are finding.

EXAMPLE 1 Finding Area under the Standard Normal Curve to the Left of a Z-Score

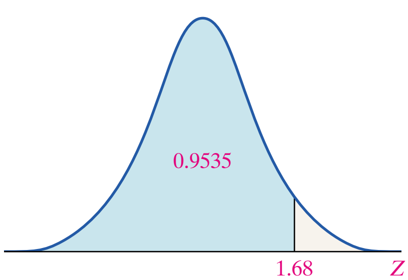
Problem: Find the area under the standard normal curve that lies to the left of $Z = 1.68$.

Approach

Step 1: Draw a standard normal curve with $Z = 1.68$ labeled, and shade the area under the curve to the left of $Z = 1.68$.

Step 2: The rows in Table IV represent the ones and tenths portion of Z, while the columns represent the hundredths portion. To find the area under the curve to the left of $Z = 1.68$, we need to split 1.68 as 1.6 and 0.08. Find the row that represents 1.6 and the column that represents 0.08 in Table IV. Identify where the row and column intersect. This value is the area.

Figure 14



Solution

Step 1: Figure 14 shows the graph of the standard normal curve with $Z = 1.68$ labeled. The area left of $Z = 1.68$ is shaded.

Step 2: A portion of Table IV is presented in Figure 15. We have enclosed the row that represents 1.6 and the column that represents 0.08. The point where the row and column intersect is the area we are seeking. The area to the left of $Z = 1.68$ is 0.9535.

Figure 15

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706

The area under a standard normal curve may also be determined using a statistical spreadsheet or graphing calculator with advanced statistical features.

EXAMPLE 2 Finding the Area under a Standard Normal Curve Using Technology

Problem: Find the area under the standard normal curve to the left of $Z = 1.68$ using a statistical spreadsheet or graphing calculator with advanced statistical features.

Approach: We will use MINITAB and a TI-84 Plus graphing calculator to find the area. The steps for determining the area under the standard normal curve for MINITAB, Excel, and the TI-83/84 Plus graphing calculators are given in the Technology Step by Step on page 344.

Result: Figure 16(a) shows the results from MINITAB, and Figure 16(b) shows the results from a TI-84 Plus graphing calculator.

Figure 16

Cumulative Distribution Function

```
Normal with mean = 0 and standard deviation = 1

      x       P( X <= x)
   1.68         0.953521
```

(a)

```
normalcdf(-1E99,
1.68,0,1)
           .9535213678
```

(b)

Notice the output for MINITAB is titled Cumulative Distribution Function. Remember, the word *cumulative* means "less than or equal to," so MINITAB is giving the area under the standard normal curve for Z less than or equal to 1.68. The command required by the TI-84 Plus is normalcdf(. The cdf stands for Cumulative Distribution Function. The TI-graphing calculators require a left and right-bound. We use $-1E99$ for the left bound to obtain areas "less than or equal to" some value.

Now Work Problem 5.

Often, rather than being interested in the area under the standard normal curve to the left of $Z = z_0$, we are interested in obtaining the area under the standard normal curve to the right of $Z = z_0$. The solution to this type of problem

In Other Words

Area right = 1 − area left

uses the fact that the area under the entire standard normal curve is 1 and the Complement Rule. Therefore,

$$\begin{pmatrix}\text{Area under the normal curve}\\ \text{to the right of } z_0\end{pmatrix} = 1 - \begin{pmatrix}\text{area to the left}\\ \text{of } z_0.\end{pmatrix}$$

$$P(Z > z_0) \quad = 1 - \quad P(Z \le z_0)$$

EXAMPLE 3 Finding Area under the Standard Normal Curve to the Right of a Z-Score

Problem: Find the area under the standard normal curve to the right of $Z = -0.46$.

Approach

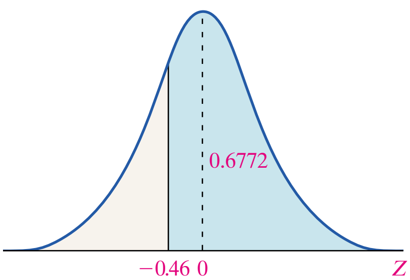
Step 1: Draw a standard normal curve with $Z = -0.46$ labeled, and shade the area under the curve to the right of $Z = -0.46$.

Step 2: Find the row that represents −0.4 and the column that represents 0.06 in Table IV. Identify where the row and column intersect. This value is the area to the *left* of $Z = -0.46$.

Step 3: The area under the standard normal curve to the right of $Z = -0.46$ is 1 minus the area to the left of $Z = -0.46$.

Solution

Figure 17



Step 1: Figure 17 shows the graph of the standard normal curve with $Z = -0.46$ labeled. The area to the right of $Z = -0.46$ is shaded.

Step 2: A portion of Table IV is presented in Figure 18. We have enclosed the row that represents −0.4 and the column that represents 0.06. The point where the row and column intersect is the area to the left of $Z = -0.46$. The area to the left of $Z = -0.46$ is 0.3228.

Figure 18

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Step 3: The area under the standard normal curve to the right of $Z = -0.46$ is 1 minus the area to the left of $Z = -0.46$.

$$\begin{aligned}\text{Area right of } -0.46 &= 1 - (\text{area left of } -0.46)\\ &= 1 - 0.3228\\ &= 0.6772\end{aligned}$$

The area to the right of $Z = -0.46$ is 0.6772.

Using Technology

Statistical spreadsheets and graphing calculators can also find areas under the standard normal curve to the right of a Z-score.

Now Work Problem 7.

The next example presents a situation in which we are interested in the area between two Z-scores.

EXAMPLE 4 Find the Area under the Standard Normal Curve between Two Z-Scores

Problem: Find the area under the standard normal curve between $Z = -1.35$ and $Z = 2.01$.

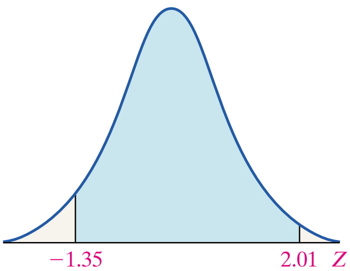
Approach

Step 1: Draw a standard normal curve with $Z = -1.35$ and $Z = 2.01$ labeled. Shade the area under the curve between $Z = -1.35$ and $Z = 2.01$.

Step 2: Find the area to the left of $Z = -1.35$. Find the area to the left of $Z = 2.01$.

Step 3: The area under the standard normal curve between $Z = -1.35$ and $Z = 2.01$ is the area to the left of $Z = 2.01$ minus the area to the left of $Z = -1.35$.

Figure 19



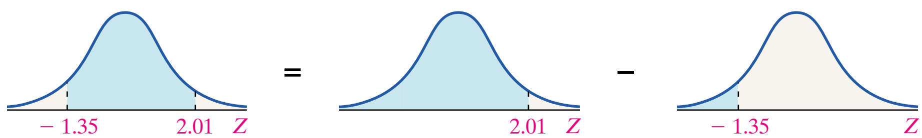
Solution

Step 1: Figure 19 shows the standard normal curve with the area between $Z = -1.35$ and $Z = 2.01$ shaded.

Step 2: Based upon Table IV, the area to the left of $Z = -1.35$ is 0.0885. The area to the left of $Z = 2.01$ is 0.9778.

Step 3: The area between $Z = -1.35$ and $Z = 2.01$ is

$$(\text{Area between } Z = -1.35 \text{ and } Z = 2.01) = (\text{area left of } Z = 2.01) - (\text{area left of } Z = -1.35)$$



$$= 0.9778 - 0.0885$$
$$= 0.8893$$

The area between $Z = -1.35$ and $Z = 2.01$ is 0.8893.

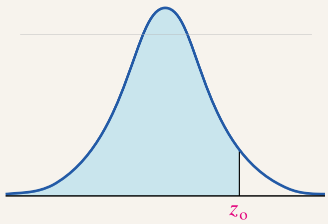
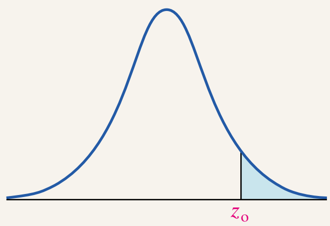
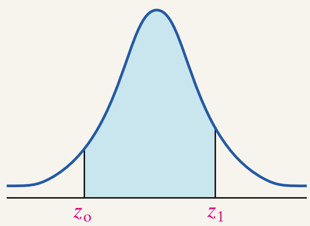
Now Work Problem 9.

We summarize the methods for obtaining area under the standard normal curve in Table 2 on page 336.

Because the normal curve extends indefinitely in both directions on the Z-axis, there is no Z-value for which the area under the curve to the left of the Z-value is 1. For example, the area to the left of $Z = 10$ is less than 1, even though graphing calculators and statistical software state that the area is 1, because they can compute a limited number of decimal places. We will follow the practice of stating the area to the left of $Z = -3.90$ or to the right of $Z = 3.90$ as <0.0001. The area under the standard normal curve to the left of $Z = 3.90$ or to the right of $Z = -3.90$ will be stated as >0.9999.

CAUTION
State the area under the standard normal curve to the left of $Z = -3.90$ as <0.0001 (not 0). State the area under the standard normal curve to the left of $Z = 3.90$ as >0.9999 (not 1).

Table 2

Problem	Approach	Solution
Find the area to the left of z_0.	Shade the area to the left of z_0. 	Use Table IV to find the row and column that correspond to z_0. The area is the value where the row and column intersect. Or use technology to find the area.
Find the area to the right of z_0.	Shade the area to the right of z_0. 	Use Table IV to find the area left of z_0. The area to the right of z_0 is 1 minus the area to the left of z_0. Or use technology to find the area.
Find the area between z_0 and z_1.	Shade the area between z_0 and z_1. 	Use Table IV to find the area to the left of z_0 and to the left of z_1. The area between z_0 and z_1 is (area to the left of z_1) − (area to the left of z_0). Or use technology to find the area.

2 Find *Z*-Scores for a Given Area

Up to this point, we have found areas given the value of a *Z*-score. Often, we are interested in finding a *Z*-score that corresponds to a given area. The procedure to follow is the reverse of the procedure for finding areas given *Z*-scores.

EXAMPLE 5 Finding a *Z*-Score from a Specified Area to the Left

Problem: Find the *Z*-score so that the area to the left of the *Z*-score is 0.32.

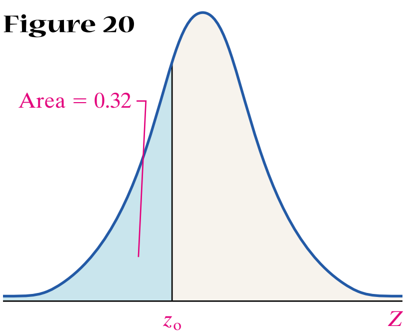
Approach

Step 1: Draw a standard normal curve with the area and corresponding unknown *Z*-score labeled.

Step 2: Look for the area in the table closest to 0.32.

Step 3: Find the *Z*-score that corresponds to the area closest to 0.32.

Figure 20



Solution

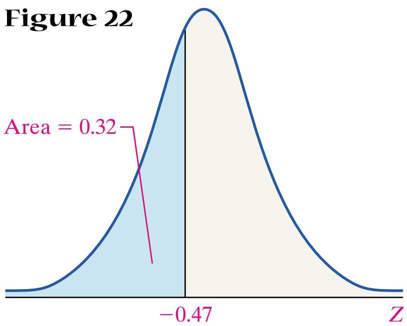
Step 1: Figure 20 shows the graph of the standard normal curve with the area of 0.32 labeled. We know z_0 must be less than 0. Do you know why?

Step 2: We refer to Table IV and look in the body of the table for an area closest to 0.32. The area closest to 0.32 is 0.3192. Figure 21 shows a partial representation of Table IV with 0.3192 labeled.

Figure 21

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4027	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859

Figure 22



Step 3: From reading the table in Figure 21, we see that the approximate *Z*-score that corresponds to an area of 0.32 to its left is −0.47. So $z_0 = -0.47$. See Figure 22.

Statistical spreadsheets or graphing calculators with advanced statistical features can also be used to determine a *Z*-score corresponding to a specified area.

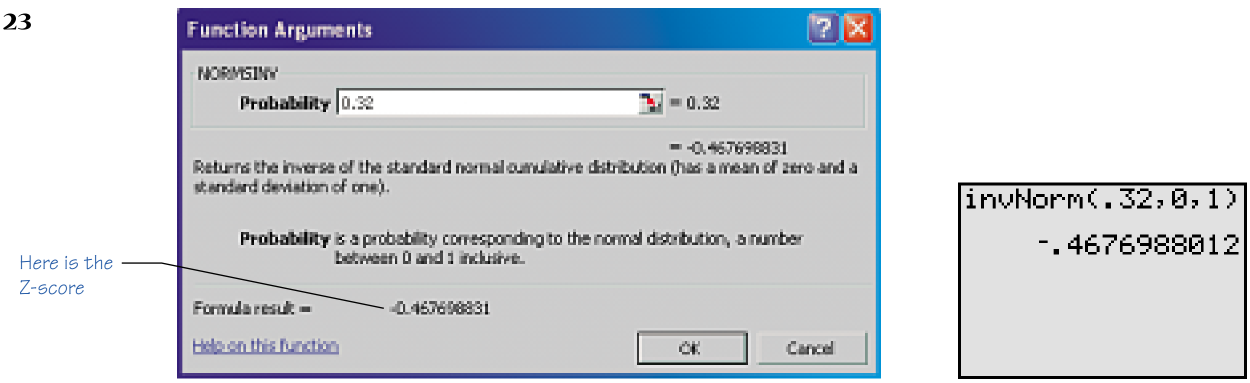
EXAMPLE 6 Finding a *Z*-Score from a Specified Area to the Left Using Technology

Problem: Find the *Z*-score such that the area to the left of the *Z*-score is 0.32 using a statistical spreadsheet or graphing calculator with advanced statistical features.

Approach: We will use Excel and a TI-84 Plus graphing calculator to find the *Z*-score. The steps for determining the *Z*-score for MINITAB, Excel, and the TI-83/84 Plus graphing calculators are given in the Technology Step by Step on page 344.

Result: Figure 23(a) shows the results from Excel, and Figure 23(b) shows the results from a TI-84 Plus graphing calculator.

Figure 23



(a) (b)

The *Z*-score that corresponds to an area of 0.32 to its left is −0.47, so $z_0 = -0.47$.

Now Work Problem 15.

It is useful to remember that if the area to the left of the *Z*-score is less than 0.5 the *Z*-score must be less than 0. If the area to the left of the *Z*-score is greater than 0.5, the *Z*-score must be greater than 0.

The next example deals with situations in which the area to the right of some unknown Z-score is given. The solution uses the fact that the area under the normal curve is 1.

EXAMPLE 7 Finding a Z-Score from a Specified Area to the Right

CAUTION

To find a Z-score given the area to the right, you must first determine the area to the left if you are using Table IV.

Problem: Find the Z-score so that the area to the right of the Z-score is 0.4332.

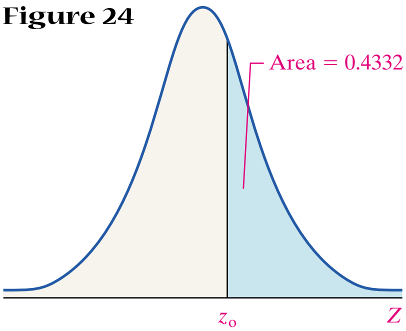
Approach

Step 1: Draw a standard normal curve with the area and corresponding unknown Z-score labeled.

Step 2: Determine the area to the left of the unknown Z-score.

Step 3: Look for the area in the table closest to the area determined in Step 2 and record the Z-score that corresponds to the closest area.

Figure 24



Solution

Step 1: Figure 24 shows the standard normal curve with the area and unknown Z-score labeled.

Step 2: Since the area under the entire normal curve is 1, the area to the left of the unknown Z-score is 1 minus the area right of the unknown Z-score. Therefore,

$$\begin{aligned}\text{Area to the left} &= 1 - \text{area to the right}\\ &= 1 - 0.4332\\ &= 0.5668\end{aligned}$$

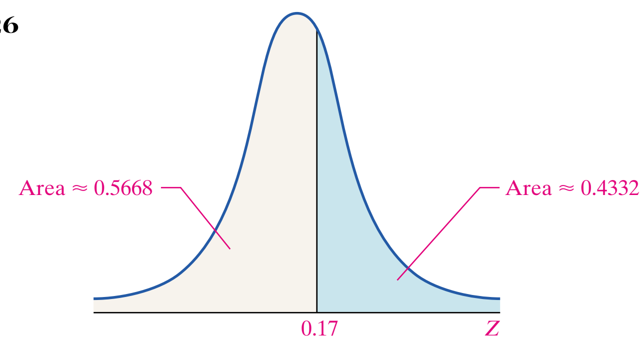
Step 3: We look in the body of Table IV for an area closest to 0.5668. See Figure 25. The area closest to 0.5668 is 0.5675.

Figure 25

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7237	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549

The approximate Z-score that corresponds to a right tail area of 0.4332 is 0.17. Therefore, $z_0 = 0.17$. See Figure 26.

Figure 26



Now Work Problem 19.

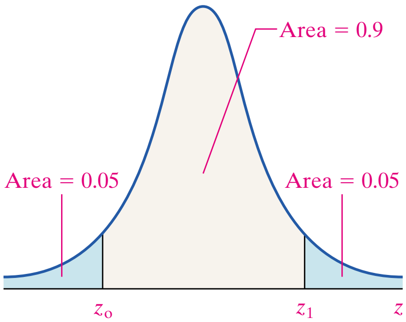
In upcoming chapters, we will be interested in finding Z-scores that separate the middle area of the standard normal curve from the area in its tails.

EXAMPLE 8 Finding the Z-Score from an Area in the Middle

Problem: Find the Z-score that divides the middle 90% of the area in the standard normal distribution from the area in the tails.

Approach

Figure 27



Step 1: Draw a standard normal curve with the middle 0.9 = 90% of the area separated from the area of 5% = 0.05 in each of the two tails. Label the unknown Z-scores z_0 and z_1.

Step 2: Look in the body of Table IV to find the area closest to 0.05.

Step 3: Determine the Z-score in the left tail.

Step 4: The area to the right of z_1 is 0.05. Therefore, the area to the left of z_1 is 0.95. Look in Table IV for an area of 0.95 and find the corresponding Z-value.

Solution

Step 1: Figure 27 shows the standard normal curve with the middle 90% of the area separated from the area in the two tails.

Step 2: We look in the body of Table IV for an area closest to 0.05. See Figure 28.

Figure 28

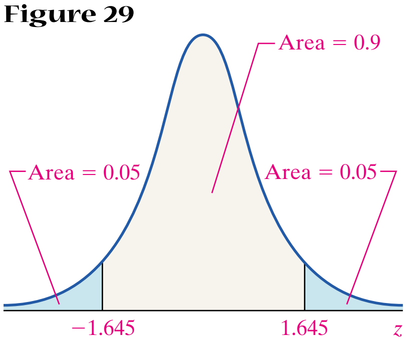
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Notice that 0.0495 and 0.0505 are equally close to 0.05. We agree to take the mean of the two Z-scores corresponding to the areas.

Step 3: The Z-score corresponding to an area of 0.0495 is −1.65. The Z-score corresponding to an area of 0.0505 is −1.64. Therefore, the approximate Z-score corresponding to an area of 0.05 to the left is

$$z_0 = \frac{-1.65 + (-1.64)}{2} = -1.645$$

Figure 29



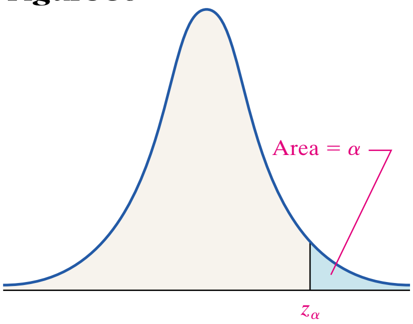
Step 4: The area to the right of z_1 is 0.05. Therefore, the area to the left of $z_1 = 1 - 0.05 = 0.95$. In Table IV, we find an area of 0.9495 corresponding to $z = 1.64$ and an area of 0.9505 corresponding to $z = 1.65$. Consequently, the approximate Z-score corresponding to an area of 0.05 to the right is

$$z_1 = \frac{1.65 + 1.64}{2} = 1.645$$

See Figure 29.

We could also obtain the solution to Example 8 using symmetry. Because the standard normal curve is symmetric about its mean, 0, the Z-score that corresponds to an area to the left of 0.05 will be the additive inverse (i.e., opposite) of the Z-score that corresponds to an area to the right of 0.05. Since the area to the left of $Z = -1.645$ is 0.05, the area to the right of $Z = 1.645$ is also 0.05.

Figure 30



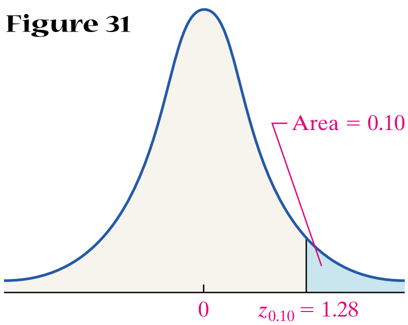
Now Work Problem 23.

We are often interested in finding the Z-score that has a specified area to the right. For this reason, we have special notation to represent this situation.

The notation z_α (pronounced "z sub alpha") is the Z-score such that the area under the standard normal curve to the right of z_α is α. Figure 30 illustrates the notation.

EXAMPLE 9 Finding the Value of z_α

Figure 31



Problem: Find the value of $z_{0.10}$.

Approach: We wish to find the Z-value such that the area under the standard normal curve to the right of the Z-value is 0.10.

Solution: The area to the right of the unknown Z-value is 0.10, so the area to the left of the Z-value is $1 - 0.10 = 0.90$. We look in Table IV for the area closest to 0.90. The area closest is 0.8997, which corresponds to a Z-value of 1.28. Therefore, $z_{0.10} = 1.28$. See Figure 31.

Now Work Problem 27.

3 Interpret the Area under the Standard Normal Curve as a Probability

Recall that the area under a normal curve can be interpreted either as a probability or as the proportion of the population with the given characteristic (as represented by an interval of numbers). When interpreting the area under the standard normal curve as a probability, we use the notation introduced in Chapter 6. For example, in Example 8, we found that the area under the standard normal curve to the left of $Z = -1.645$ is 0.05; therefore, the probability of randomly selecting a standard normal random variable that is less than -1.645 is 0.05. We write this statement with the notation $P(Z < -1.645) = 0.05$.

We will use the following notation to denote probabilities of a standard normal random variable, Z.

Notation for the Probability of a Standard Normal Random Variable

$P(a < Z < b)$	represents the probability a standard normal random variable is between a and b.
$P(Z > a)$	represents the probability a standard normal random variable is greater than a.
$P(Z < a)$	represents the probability a standard normal random variable is less than a.

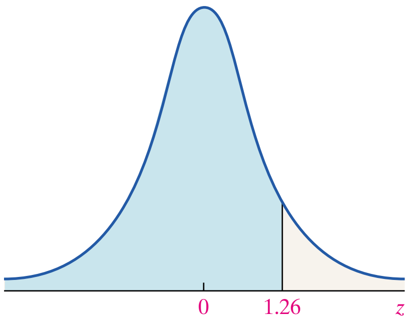
EXAMPLE 10 Finding Probabilities of Standard Normal Random Variables

Problem: Evaluate $P(Z < 1.26)$.

Approach

Step 1: Draw a standard normal curve with the area we desire shaded.

Figure 32



Step 2: Use Table IV to find the area of the shaded region. This area represents the probability.

Solution

Step 1: Figure 32 shows the standard normal curve with the area to the left of $Z = 1.26$ shaded.

Step 2: Using Table IV, we find the area under the standard normal curve to the left of $Z = 1.26$ is 0.8962. Therefore, $P(Z < 1.26) = 0.8962$.

Now Work Problem 33.

For any continuous random variable, the probability of observing a specific value of the random variable is 0. For example, for a standard normal random variable, $P(a) = 0$ for any value of a. This is because there is no area under the standard normal curve associated with a single value, so the probability must be 0. So, the following probabilities are equivalent.

$$P(a < Z < b) = P(a \leq Z < b) = P(a < Z \leq b) = P(a \leq Z \leq b)$$

For example, $P(Z < 1.26) = P(Z \leq 1.26) = 0.8962$.

7.2 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. State the properties of the standard normal curve.

2. If the area under the standard normal curve to the left of $Z = 1.20$ is 0.8849, what is the area under the standard normal curve to the right of $Z = 1.20$?

3. *True or False*: The area under the standard normal curve to the left of $Z = 5.30$ is 1. Support your answer.

4. Explain why $P(Z < -1.30) = P(Z \leq -1.30)$.

Skill Building

In Problems 5–12, find the indicated areas. For each problem, be sure to draw a standard normal curve and shade the area that is to be found.

5. NW Determine the area under the standard normal curve that lies to the left of
(a) $Z = -2.45$
(b) $Z = -0.43$
(c) $Z = 1.35$
(d) $Z = 3.49$

6. Determine the area under the standard normal curve that lies to the left of
(a) $Z = -3.49$
(b) $Z = -1.99$
(c) $Z = 0.92$
(d) $Z = 2.90$

7. NW Determine the area under the standard normal curve that lies to the right of
(a) $Z = -3.01$
(b) $Z = -1.59$
(c) $Z = 1.78$
(d) $Z = 3.11$

8. Determine the area under the standard normal curve that lies to the right of
(a) $Z = -3.49$
(b) $Z = -0.55$
(c) $Z = 2.23$
(d) $Z = 3.45$

9. NW Determine the area under the standard normal curve that lies between
(a) $Z = -2.04$ and $Z = 2.04$
(b) $Z = -0.55$ and $Z = 0$
(c) $Z = -1.04$ and $Z = 2.76$

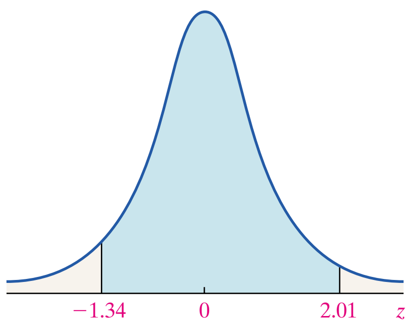
10. Determine the area under the standard normal curve that lies between
(a) $Z = -2.55$ and $Z = 2.55$
(b) $Z = -1.67$ and $Z = 0$
(c) $Z = -3.03$ and $Z = 1.98$

11. Determine the area under the standard normal curve
(a) to the left of $Z = -2$ or to the right of $Z = 2$
(b) to the left of $Z = -1.56$ or to the right of $Z = 2.56$
(c) to the left of $Z = -0.24$ or to the right of $Z = 1.20$

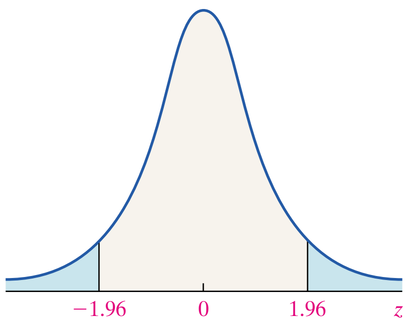
12. Determine the area under the standard normal curve
(a) to the left of $Z = -2.94$ or to the right of $Z = 2.94$
(b) to the left of $Z = -1.68$ or to the right of $Z = 3.05$
(c) to the left of $Z = -0.88$ or to the right of $Z = 1.23$

In Problems 13 and 14, find the area of the shaded region for each standard normal curve.

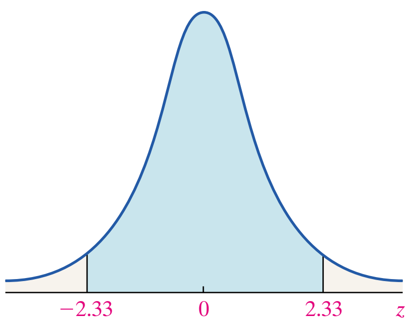
13. (a)



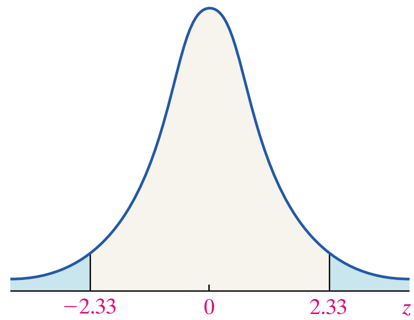
(b)



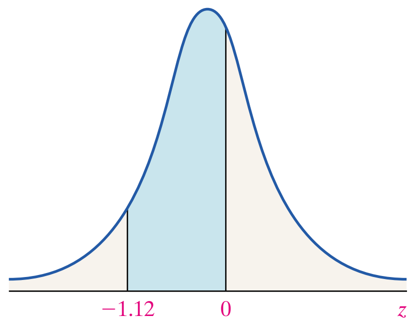
(c)



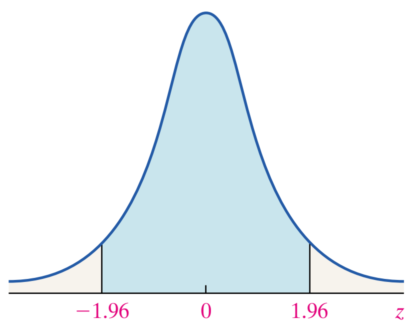
14. (a)



(b)



(c)



In Problems 15–26, find the indicated Z-score. Be sure to draw a standard normal curve that depicts the solution.

15. NW Find the Z-score such that the area under the standard normal curve to the left is 0.1.

16. Find the Z-score such that the area under the standard normal curve to the left is 0.2.

17. Find the Z-score such that the area under the standard normal curve to the left is 0.98.

18. Find the Z-score such that the area under the standard normal curve to the left is 0.85.

19. NW Find the Z-score such that the area under the standard normal curve to the right is 0.25.

20. Find the Z-score such that the area under the standard normal curve to the right is 0.35.

21. Find the Z-score such that the area under the standard normal curve to the right is 0.89.

22. Find the Z-score such that the area under the standard normal curve to the right is 0.75.

23. NW Find the Z-scores that separate the middle 80% of the distribution from the area in the tails of the standard normal distribution.

24. Find the Z-scores that separate the middle 70% of the distribution from the area in the tails of the standard normal distribution.

25. Find the Z-scores that separate the middle 99% of the distribution from the area in the tails of the standard normal distribution.

26. Find the Z-scores that separate the middle 94% of the distribution from the area in the tails of the standard normal distribution.

In Problems 27–32, find the value of z_α.

27. NW $z_{0.05}$

28. $z_{0.35}$

29. $z_{0.01}$

30. $z_{0.02}$

31. $z_{0.20}$

32. $z_{0.15}$

In Problems 33–44, find the indicated probability of the standard normal random variable Z.

33. $P(Z < 1.93)$

34. $P(Z < -0.61)$

35. $P(Z > -2.98)$

36. $P(Z > 0.92)$

37. $P(-1.20 \leq Z < 2.34)$

38. $P(1.23 < Z \leq 1.56)$

39. $P(Z \geq 1.84)$

40. $P(Z \geq -0.92)$

41. $P(Z \leq 0.72)$

42. $P(Z \leq -2.69)$

43. $P(Z < -2.56 \text{ or } Z > 1.39)$

44. $P(Z < -0.38 \text{ or } Z > 1.93)$

Applying the Concepts

45. The Empirical Rule The Empirical Rule states that about 68% of the data in a bell-shaped distribution lies within 1 standard deviation of the mean. This means about 68% of the data lie between $Z = -1$ and $Z = 1$. Verify this result. Verify that about 95% of the data lie within 2 standard deviations of the mean. Finally, verify that about 99.7% of the data lie within 3 standard deviations of the mean.

46. According to Table IV, the area under the standard normal curve to the left of $Z = -1.34$ is 0.0901. Without consulting Table IV, determine the area under the standard normal curve to the right of $Z = 1.34$.

47. According to Table IV, the area under the standard normal curve to the left of $Z = -2.55$ is 0.0054. Without consulting Table IV, determine the area under the standard normal curve to the right of $Z = 2.55$.

48. According to Table IV, the area under the standard normal curve between $Z = -1.50$ and $Z = 0$ is 0.4332. Without consulting Table IV, determine the area under the standard normal curve between $Z = 0$ and $Z = 1.50$.

49. According to Table IV, the area under the standard normal curve between $Z = -1.24$ and $Z = -0.53$ is 0.1906. Without consulting Table IV, determine the area under the standard normal curve between $Z = 0.53$ and $Z = 1.24$.

50. (a) Suppose $P(Z < a) = 0.9938$; find a.

(b) Suppose $P(Z \geq a) = 0.4404$; find a.

(c) Suppose $P(-b < Z < b) = 0.8740$; find b.

Technology Step by Step — The Standard Normal Distribution

TI-83/84 Plus

Finding Areas under the Standard Normal Curve

***Step 1*:** From the HOME screen, press 2nd VARS to access the DISTRibution menu.

***Step 2*:** Select `2:normalcdf(`

***Step 3*:** With `normalcdf(` on the HOME screen, type *lowerbound, upperbound, 0, 1*). For example, to find the area left of $Z = 1.26$ under the standard normal curve, type

```
Normalcdf(-1E99, 1.26, 0, 1)
```

and hit ENTER.

***Note*:** When there is no lowerbound, enter −1E99. When there is no upperbound, enter 1E99. The E shown is scientific notation; it is [2nd] [,] on the keyboard.

Finding *Z*-Scores Corresponding to an Area

***Step 1*:** From the HOME screen, press 2nd VARS to access the DISTRibution menu.

***Step 2*:** Select `3:invNorm(`.

***Step 3*:** With `invNorm(` on the HOME screen, type *"area left", 0, 1*). For example, to find the *Z*-score such that the area under the normal curve left of the *Z*-score is 0.79, type

```
InvNorm(0.79, 0 , 1)
```

and hit ENTER.

MINITAB

Finding Areas under the Standard Normal Curve

***Step 1*:** MINITAB will find an area to the left of a specified *Z*-score. Select the **Calc** menu, highlight **Probability Distributions**, and highlight **Normal** . . . .

***Step 2*:** Select **Cumulative Probability**. Set the mean to 0 and the standard deviation to 1. Select **Input Constant**, and enter the specified *Z*-score. Click OK.

Finding *Z*-Scores Corresponding to an Area

***Step 1*:** MINITAB will find the *Z*-score for an area to the left of an unknown *Z*-score. Select the **Calc** menu, highlight **Probability Distributions**, and highlight **Normal** . . . .

***Step 2*:** Select **Inverse Cumulative Probability**. Set the mean to 0 and the standard deviation to 1. Select **Input Constant**, and enter the specified area. Click OK.

Excel

Finding Areas under the Standard Normal Curve

Step 1*:** Excel will find the area to the left of a specified *Z*-score. Select the ***fx button from the tool bar. In **Function Category:**, select "Statistical." In **Function Name:**, select "NormsDist." Click OK.

***Step 2*:** Enter the specified *Z*-score. Click OK.

Finding *Z*-Scores Corresponding to an Area

Step 1*:** Excel will find the *Z*-score for an area to the left of an unknown *Z*-score. Select the ***fx button from the tool bar. In **Function Category:**, select "Statistical." In **Function Name:**, select "NormsInv." Click OK.

***Step 2*:** Enter the specified area. Click OK.

7.3 Applications of the Normal Distribution

Preparing for This Section Before getting started, review the following:

- Percentiles (Section 3.4, pp. 151–153)

Objectives

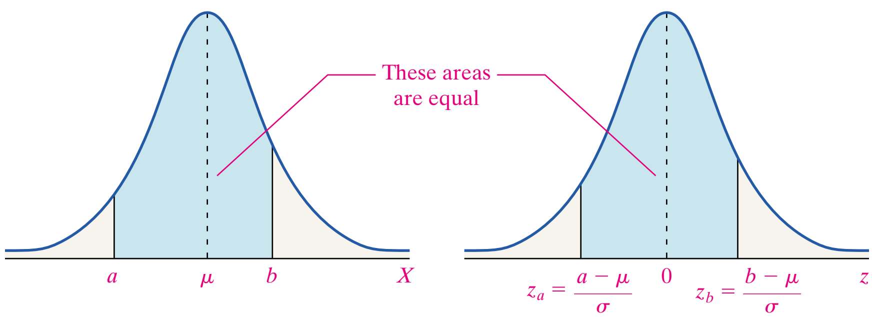
1. **Find and interpret the area under a normal curve**
2. **Find the value of a normal random variable**

1 Find and Interpret the Area under a Normal Curve

Suppose that a random variable X is normally distributed with mean μ and standard deviation σ. The area below the normal curve represents a proportion or probability.

From the discussions in Section 7.1, we know that finding the area under a normal curve requires that we transform a normal random variable X with mean μ and standard deviation σ into a standard normal random variable Z with mean 0 and standard deviation 1. This is accomplished by letting $Z = \dfrac{X - \mu}{\sigma}$ and using Table IV to find the area under the standard normal curve. This idea is illustrated in Figure 33.

Figure 33



Now that we have the ability to find the area under a standard normal curve, we can find the area under any normal curve. We summarize the procedure next.

Finding the Area under Any Normal Curve

***Step 1*:** Draw a normal curve and shade the desired area.

***Step 2*:** Convert the values of X to Z-scores using $Z = \dfrac{X - \mu}{\sigma}$.

***Step 3*:** Draw a standard normal curve and shade the area desired.

***Step 4*:** Find the area under the standard normal curve. This area is equal to the area under the normal curve drawn in Step 1.

EXAMPLE 1 **Finding Area under a Normal Curve**

***Problem*:** A pediatrician obtains the heights of her 200 three-year-old female patients. The heights are approximately normally distributed, with mean 38.72 inches and standard deviation 3.17 inches. Use the normal model to determine the proportion of the 3-year-old females that have a height less than 35 inches.

***Approach*:** Follow Steps 1 through 4.

Solution

***Step 1*:** Figure 34 shows the normal curve with the area to the left of 35 shaded.

***Step 2*:** We convert $X = 35$ to a standard normal random variable Z.

$$Z = \frac{X - \mu}{\sigma} = \frac{35 - 38.72}{3.17} = -1.17$$

***Step 3*:** Figure 35 shows the standard normal curve with the area to the left of $Z = -1.17$ shaded. The area to the left of $Z = -1.17$ is equal to the area to the left of $X = 35$.

Figure 34

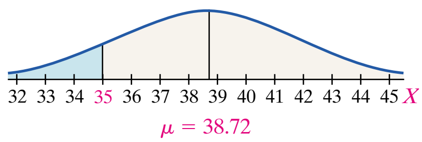
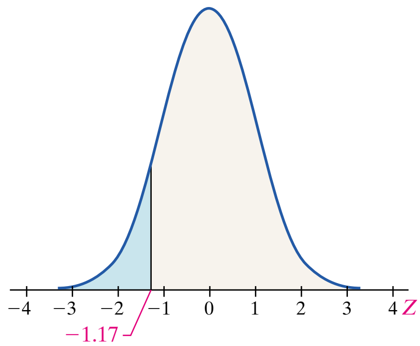


Figure 35



***Step 4*:** Using Table IV, we find the area to the left of $Z = -1.17$ is 0.1210. The normal model indicates that the proportion of the pediatrician's 3-year-old females that are less than 35 inches tall is 0.1210.

Table 3

Height (inches)	Relative Frequency
29.0–29.9	0.005
30.0–30.9	0.005
31.0–31.9	0.005
32.0–32.9	0.025
33.0–33.9	0.02
34.0–34.9	0.055
35.0–35.9	0.075
36.0–36.9	0.09
37.0–37.9	0.115
38.0–38.9	0.15
39.0–39.9	0.12
40.0–40.9	0.11
41.0–41.9	0.07
42.0–42.9	0.06
43.0–43.9	0.035
44.0–44.9	0.025
45.0–45.9	0.025
46.0–46.9	0.005
47.0–47.9	0.005

According to the results of Example 1, the probability that a randomly selected 3-year-old female is shorter than 35 inches is 0.1210. If the normal curve is a good model for determining probabilities (or proportions), then about 12.1% of the 200 three-year-olds in Table 1 should be shorter than 35 inches. For convenience, the information provided in Table 1 is repeated in Table 3.

From the relative frequency distribution in Table 3, we determine that $0.005 + 0.005 + 0.005 + 0.025 + 0.02 + 0.055 = 0.115 = 11.5\%$ of the 3-year-olds are less than 35 inches tall. The results based on the normal curve are in close agreement with the actual results. The normal curve accurately models the heights of 3-year-old females.

Because the area under the normal curve represents a proportion, we can also use the area under the normal curve to find percentile ranks of scores. Recall that the kth percentile divides the lower $k\%$ of a data set from the upper $(100 - k)\%$. In Example 1, 12% of the females have a height less than 35 inches, and 88% of the females have a height greater than 35 inches. Therefore, a child whose height is 35 inches is at the 12th percentile.

Statistical software and graphing calculators with advanced statistical features can also be used to find areas under any normal curve.

EXAMPLE 2 Finding Area under a Normal Curve Using Technology

***Problem*:** Find the percentile rank of a 3-year-old female whose height is 43 inches using a statistical spreadsheet or graphing calculator with advanced statistical features. From Example 1, we know the heights are approximately normally distributed with a mean of 38.72 inches and standard deviation of 3.17 inches.

***Approach*:** We will use a TI-84 Plus graphing calculator to find the area. The steps for determining the area under the standard normal curve for MINITAB,

Figure 36

Excel, and the TI-83/84 Plus graphing calculators are given in the Technology Step by Step on page 353.

Result: Figure 36 shows the results from a TI-84 Plus graphing calculator. The area under the normal curve to the left of 43 is 0.91. Therefore, 91% of the heights are less than 43 inches and 9% of the heights are more than 43 inches. A child whose height is 43 inches is at the 91st percentile.

EXAMPLE 3 Finding the Probability of a Normal Random Variable

Problem: For the pediatrician presented in Example 1, use the normal distribution to compute the probability that a randomly selected 3-year-old female is between 35 and 40 inches tall, inclusive. That is, find $P(35 \le X \le 40)$.

Approach: We follow the Steps 1 through 4 on page 345.

Solution

Step 1: Figure 37 shows the normal curve with the area between $X_1 = 35$ and $X_2 = 40$ shaded.

Step 2: Convert the values of $X_1 = 35$ and $X_2 = 40$ to Z-scores.

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{35 - 38.72}{3.17} = -1.17$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{40 - 38.72}{3.17} = 0.40$$

Step 3: Figure 38 shows the graph of the standard normal curve with the area between $Z_1 = -1.17$ and $Z_2 = 0.40$ shaded.

Figure 37

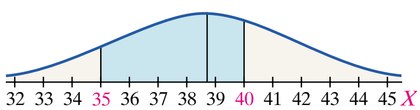
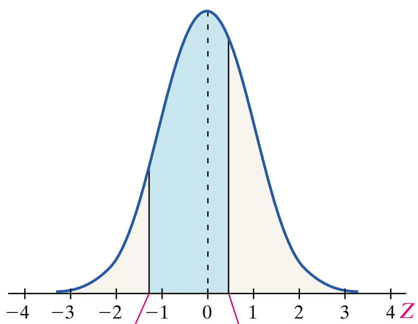


Figure 38



Step 4: Using Table IV, we find that the area to the left of $Z_2 = 0.40$ is 0.6554 and the area to the left of $Z_1 = -1.17$ is 0.1210. Therefore, the area between $Z_1 = -1.17$ and $Z_2 = 0.40$ is $0.6554 - 0.1210 = 0.5344$. We conclude that the probability a randomly selected 3-year-old female is between 35 and 40 inches tall is 0.5344. That is, $P(35 \le X \le 40) = P(-1.17 \le Z \le 0.40) = 0.5344$. If we randomly selected one 3-year-old female 100 times, we would expect to select a child who is between 35 and 40 inches tall about 53 times.

Now Work Problem 19.

In Other Words
The normal probability density function is used to model random variables that appear to be normal (such as girls' heights). A good model is one that yields results that are close to reality.

According to the relative frequency distribution in Table 3, the proportion of the 200 three-year-old females with heights between 35 inches and 40 inches is $0.075 + 0.09 + 0.115 + 0.15 + 0.12 = 0.55 = 55\%$. This is very close to the probability obtained in Example 3!

2 Find the Value of a Normal Random Variable

Often, rather than being interested in the proportion or probability of a normal random variable, we are interested in calculating the value of a normal random variable required for the variable to correspond to a certain proportion or probability. For example, we might want to know the height of a 3-year-old girl at the 20th percentile. This means we want to know the height of a 3-year-old girl who is taller than 20% of all 3-year-old girls.

Procedure for Finding the Value of a Normal Random Variable Corresponding to a Specified Proportion, Probability, or Percentile

***Step 1*:** Draw a normal curve and shade the area corresponding to the proportion, probability, or percentile.

***Step 2*:** Use Table IV to find the Z-score that corresponds to the shaded area.

***Step 3*:** Obtain the normal value from the fact that $X = \mu + Z\sigma$.*

EXAMPLE 4 Finding the Value of a Normal Random Variable

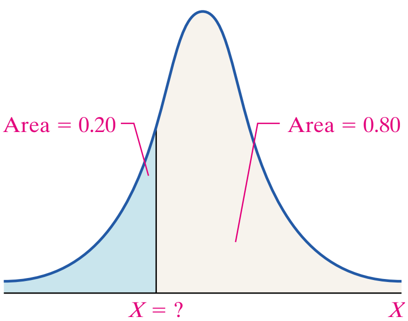
***Problem*:** The heights of a pediatrician's 200 three-year-old females are approximately normally distributed with mean 38.72 inches and standard deviation 3.17 inches. Find the height of a 3-year-old female at the 20th percentile. That is, find the height of a 3-year-old female that separates the bottom 20% from the top 80%.

***Approach*:** We follow Steps 1 through 3.

Solution

***Step 1*:** Figure 39 shows the normal curve with the unknown value of X separating the bottom 20% of the distribution from the top 80% of the distribution.

Figure 39



***Step 2*:** From Table IV, the area closest to 0.20 is 0.2005. The corresponding Z-score is -0.84.

***Step 3*:** The height of a 3-year-old female that separates the bottom 20% of the data from the top 80% is computed as follows:

$$\begin{aligned} X &= \mu + Z\sigma \\ &= 38.72 + (-0.84)(3.17) \\ &= 36.1 \text{ inches} \end{aligned}$$

The height of a 3-year-old female at the 20th percentile is 36.1 inches.

EXAMPLE 5 Finding the Value of a Normal Random Variable Using Technology

***Problem*:** Use a statistical spreadsheet or a graphing calculator with advanced statistical features to verify the results of Example 4. That is, find the height for a 3-year-old female that is at the 20th percentile, assuming females' heights are approximately normally distributed with a mean of 38.72 inches and a standard deviation of 3.17 inches.

***Approach*:** We will use MINITAB to find the height at the 20th percentile. The steps for determining the area under the standard normal curve for

*$Z = \dfrac{X - \mu}{\sigma}$ Formula for standardizing a random variable X

$Z\sigma = X - \mu$ Multiply both sides by σ.

$X = \mu + Z\sigma$ Add μ to both sides.

MINITAB, Excel, and the TI-83/84 Plus graphing calculators are given in the Technology Step by Step on page 353.

Result: Figure 40 shows the results obtained from MINITAB. The height of a three-year-old female at the 20th percentile is 36.1 inches.

Figure 40

Inverse Cumulative Distribution Function

```
Normal with mean = 38.7200 and standard deviation = 3.17000

P ( x <= x)        x
    0.2000     36.0521
```

Now Work Problem 27(a).

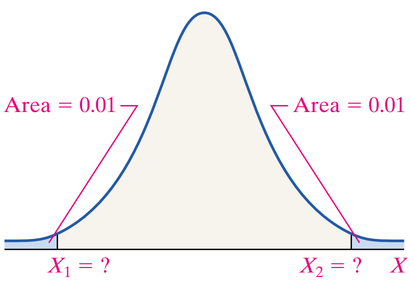
EXAMPLE 6 Finding the Value of a Normal Random Variable

Problem: The heights of a pediatrician's 200 three-year-old females are approximately normally distributed with mean 38.72 inches and standard deviation 3.17 inches. The pediatrician wishes to determine the heights that separate the middle 98% of the distribution from the bottom 1% and top 1%. In other words, find the 1st and 99th percentiles.

Approach: We follow Steps 1 through 3 given on page 348.

Solution

Figure 41



***Step 1*:** Figure 41 shows the normal curve with the unknown values of X separating the bottom and top 1% of the distribution from the middle 98% of the distribution.

***Step 2*:** First, we will find the Z-score that corresponds to the area 0.01 to the left. From Table IV, the area closest to 0.01 is 0.0099. The corresponding Z-score is -2.33. The Z-score that corresponds to the area 0.01 to the right is the Z-score that has the area 0.99 to the left. The area closest to 0.99 is 0.9901. The corresponding Z-score is 2.33.

***Step 3*:** The height of a 3-year-old female that separates the bottom 1% of the distribution from the top 99% is

$$\begin{aligned} X_1 &= \mu + Z\sigma \\ &= 38.72 + (-2.33)(3.17) \\ &= 31.3 \text{ inches} \end{aligned}$$

The height of a 3-year-old female that separates the top 1% of the distribution from the bottom 99% is

$$\begin{aligned} X_2 &= \mu + Z\sigma \\ &= 38.72 + (2.33)(3.17) \\ &= 46.1 \text{ inches} \end{aligned}$$

A 3-year-old female whose height is less than 31.3 inches is in the bottom 1% of all 3-year-old females, and a 3-year-old female whose height is more than 46.1 inches is in the top 1% of all 3-year-old females. The pediatrician might use this information to identify those patients who have unusual heights.

Now Work Problem 27(b).

CAUTION

If you are given a value of the random variable and asked to find the probability, proportion, or percentile corresponding to the random variable, convert the random variable to a Z-score using $Z = \frac{X - \mu}{\sigma}$ and find the area from the table. If you are using a TI-graphing calculator, use normalcdf.

If you are asked to find the random variable corresponding to a probability, proportion, or percentile, find the area that represents the given probability, proportion, or percentile and use $XZ = \mu + \sigma$ to find the value of the random variable. If you are using a TI-graphing calculator, use invNorm.

7.3 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. Describe the procedure for finding the area under any normal curve.
2. Describe the procedure for finding the score corresponding to a probability.

Skill Building

In Problems 3–12, assume the random variable X is normally distributed with mean $\mu = 50$ and standard deviation $\sigma = 7$. Compute the following probabilities. Be sure to draw a normal curve with the area corresponding to the probability shaded.

3. $P(X > 35)$
4. $P(X > 65)$
5. $P(X \le 45)$
6. $P(X \le 58)$
7. $P(40 < X < 65)$
8. $P(56 < X < 68)$
9. $P(55 \le X \le 70)$
10. $P(40 \le X \le 49)$
11. $P(38 < X \le 55)$
12. $P(56 \le X < 66)$

In Problems 13–16, assume the random variable X, is normally distributed with mean $\mu = 50$ and standard deviation $\sigma = 7$. Find each indicated percentile for X.

13. The 9th percentile
14. The 90th percentile
15. The 81st percentile
16. The 38th percentile

Applying the Concepts

17. **Egg Incubation Times** The mean incubation time of fertilized chicken eggs kept at 100.5°F in a still-air incubator is 21 days. Suppose the incubation times are approximately normally distributed with a standard deviation of 1 day. (*Source*: University of Illinois Extension.)
 (a) What is the probability that a randomly selected fertilized chicken egg hatches in less than 20 days?
 (b) What is the probability that a randomly selected fertilized chicken egg takes over 22 days to hatch?
 (c) What is the probability that a randomly selected fertilized chicken egg hatches between 19 and 21 days?
 (d) Would it be unusual for an egg to hatch in less than 18 days?

18. **Medical Residents** In a 2003 study, the Accreditation Council for Graduate Medical Education found that medical residents' mean number of hours worked in a week is 81.7. Suppose the number of hours worked per week by medical residents is approximately normally distributed with a standard deviation of 6.9 hours.
 (*Source*: www.medrecinst.com)
 (a) What is the probability that a randomly selected medical resident works more than 80 hours in a week?
 (b) What is the probability that a randomly selected medical resident works more than 100 hours in a week?
 (c) What is the probability that a randomly selected medical resident works less than 60 hours in a week?
 (d) Would it be unusual for a medical resident to work less than 70 hours in a week?

19. **Chips Ahoy! Cookies** NW The number of chocolate chips in an 18-ounce bag of Chips Ahoy! chocolate chip cookies is approximately normally distributed with a mean of 1262 chips and standard deviation 118 chips according to a study by cadets of the U.S. Air Force Academy.
 (*Source*: Brad Warner and Jim Rutledge, *Chance*, Vol. 12, No. 1, 1999, pp. 10–14.)
 (a) What is the probability that a randomly selected 18-ounce bag of Chips Ahoy! cookies contains between 1000 and 1400 chocolate chips?
 (b) What is the probability that a randomly selected 18-ounce bag of Chips Ahoy! cookies contains fewer than 1000 chocolate chips?
 (c) What proportion of 18-ounce bags of Chip Ahoy! cookies contains more than 1200 chocolate chips?
 (d) What proportion of 18-ounce bags of Chip Ahoy! cookies contains fewer than 1125 chocolate chips?
 (e) What is the percentile rank of an 18-ounce bag of Chip Ahoy! cookies that contains 1475 chocolate chips?
 (f) What is the percentile rank of an 18-ounce bag of Chip Ahoy! cookies that contains 1050 chocolate chips?

20. **Earthquakes** The magnitude of earthquakes since 1900 that measure 0.1 or higher on the Richter scale in California is approximately normally distributed, with $\mu = 6.2$ and $\sigma = 0.5$, according to data obtained from the U.S. Geological Survey.

(a) What is the probability that a randomly selected earthquake in California has a magnitude of 6.0 or higher?
(b) What is the probability that a randomly selected earthquake in California has a magnitude less than 6.4?
(c) What is the probability that a randomly selected earthquake in California has a magnitude between 5.8 and 7.1?
(d) The great San Francisco Earthquake of 1906 had a magnitude of 8.25. Is an earthquake of this magnitude unusual in California?
(e) What is the percentile rank of a California earthquake that measures 6.8 on the Richter scale?
(f) What is the percentile rank of a California earthquake that measures 5.1 on the Richter scale?

21. Hybrid Car Introduced in the 2000 model year, the Honda Insight was the first hybrid automobile sold in the United States. The mean gas mileage for the model year 2005 Insight with an automatic transmission is 56 miles per gallon on the highway. Suppose the gasoline mileage of this automobile is approximately normally distributed with a standard deviation of 3.2 miles per gallon. (*Source*: www.fueleconomy.gov)
(a) What proportion of 2005 Honda Insights with automatic transmission gets over 60 miles per gallon on the highway?
(b) What proportion of 2005 Honda Insights with automatic transmission gets 50 miles per gallon or less on the highway?
(c) What proportion of 2005 Honda Insights with automatic transmission gets between 58 and 62 miles per gallon on the highway?
(d) What is the probability that a randomly selected 2005 Honda Insight with an automatic transmission gets less than 45 miles per gallon on the highway?

22. Light Bulbs General Electric manufactures a decorative Crystal Clear 60-watt light bulb that it advertises will last 1500 hours. Suppose the lifetimes of the light bulbs are approximately normally distributed with a mean of 1550 hours and a standard deviation of 57 hours.
(a) What proportion of the light bulbs will last less than the advertised time?
(b) What proportion of the light bulbs will last more than 1650 hours?
(c) What is the probability that a randomly selected GE Crystal Clear 60-watt light bulb lasts between 1625 and 1725 hours?
(d) What is the probability that a randomly selected GE Crystal Clear 60-watt light bulb lasts longer than 1400 hours?

23. Heights of Females As reported by the U.S. National Center for Health Statistics, the mean height of females 20 to 29 years old is $\mu = 64.1$ inches. If height is approximately normally distributed with $\sigma = 2.8$ inches, answer the following questions:
(a) What is the percentile rank of a 20- to 29-year-old female who is 60 inches tall?
(b) What is the percentile rank of a 20- to 29-year-old female who is 70 inches tall?
(c) What proportion of 20- to 29-year-old females are between 60 and 70 inches tall?
(d) Would it be unusual for a 20- to 29-year-old female to be taller than 70 inches?

24. Gestation Period The length of human pregnancies are approximately normally distributed with mean $\mu = 266$ days and standard deviation $\sigma = 16$ days.
(a) What percent of pregnancies lasts more than 270 days?
(b) What percent of pregnancies lasts less than 250 days?
(c) What percent of pregnancies lasts between 240 and 280 days?
(d) What is the probability that a randomly selected pregnancy lasts more than 280 days?
(e) What is the probability that a randomly selected pregnancy lasts no more than 245 days?
(f) A "very preterm" baby is one whose gestation period is less than 224 days. What proportion of births is "very preterm"?

25. Manufacturing Steel rods are manufactured with a mean length of 25 centimeter (cm). Because of variability in the maufacturing process, the lengths of the rods are approximately normally distributed with a standard deviation of 0.07 cm.
(a) What proportion of rods has a length less than 24.9 cm?
(b) Any rods that are shorter than 24.85 cm or longer than 25.15 cm are discarded. What proportion of rods will be discarded?
(c) Using the results of part (b), if 5000 rods are manufactured in a day, how many should the plant manager expect to discard?
(d) If an order comes in for 10,000 steel rods, how many rods should the plant manager manufacture if the order states that all rods must be between 24.9 cm and 25.1 cm?

26. Manufacturing Ball bearings are manufactured with a mean diameter of 5 millimeter (mm). Because of variability in the manufacturing process, the diameters of the ball bearings are approximately normally distributed with a standard deviation of 0.02 mm.
(a) What proportion of ball bearings has a diameter more than 5.03 mm?
(b) Any ball bearings that have a diameter less than 4.95 mm or greater than 5.05 mm are discarded. What proportion of ball bearings will be discarded?
(c) Using the results of part (b), if 30,000 ball bearings are manufactured in a day, how many should the plant manager expect to discard?
(d) If an order comes in for 50,000 ball bearings, how many bearings should the plant manager manufacture if the order states that all ball bearings must be between 4.97 mm and 5.03 mm?

27. Egg Incubation Times The mean of the incubation time of fertilized chicken eggs kept at 100.5°F in a still-air NW incubator is 21 days. Suppose the incubation times are approximately normally distributed with a standard deviation of 1 day.

(*Source*: University of Illinois Extension.)

(a) Determine the 17th percentile for incubation times of fertilized chicken eggs.
(b) Determine the incubation times that make up the middle 95% of fertilized chicken eggs?

28. Medical Residents In a 2003 study, the Accreditation Council for Graduate Medical Education found that medical residents' mean number of hours worked in a week is 81.7. Suppose the number of hours worked per week by medical residents is approximately normally distributed with a standard deviation of 6.9 hours.

(*Source*: www.medrecinst.com)

(a) Determine the 75th percentile for the number of hours worked in a week by medical residents.
(b) Determine the number of hours worked in a week that makes up the middle 80% of medical residents.

29. Chips Ahoy! Cookies The number of chocolate chips in an 18-ounce bag of Chips Ahoy! chocolate chip cookies is approximately normally distributed with a mean of 1262 chips and a standard deviation of 118 chips, according to a study by cadets of the U.S. Air Force Academy.

(*Source*: Brad Warner and Jim Rutledge, *Chance*, Vol. 12, No. 1, 1999, pp. 10–14.)

(a) Determine the 30th percentile for the number of chocolate chips in an 18-ounce bag of Chips Ahoy! cookies.
(b) Determine the number of chocolate chips in a bag of Chips Ahoy! that make up the middle 99% of bags.

30. Earthquakes The magnitude of earthquakes since 1900 that measure 0.1 or higher on the Richter scale in California is approximately normally distributed with $\mu = 6.2$ and $\sigma = 0.5$, according to data obtained from the U.S. Geological Survey.

(a) Determine the 40th percentile of the magnitude of earthquakes in California.
(b) Determine the magnitude of earthquakes that make up the middle 85% of magnitudes.

31. Hybrid Car Introduced in the 2000 model year, the Honda Insight was the first hybrid automobile sold in the United States. The mean was mileage for the model year 2005 Insight with an automatic transmission is 56 miles per gallon on the highway. Suppose the gasoline mileages of these automobiles are approximately normally distributed with standard deviation 3.2 miles per gallon.

(*Source*: www.fuel.economy.gov)

(a) Determine the 97th percentile gasoline mileage for the model year 2005 Insight with an automatic transmission.
(b) Determine the mileage that makes up the middle 86% gasoline mileage of model year 2005 Insight with an automatic transmission.

32. Speedy Lube The time required for Speedy Lube to complete an oil change service on an automobile approximately follows a normal distribution, with a mean of 17 minutes and a standard deviation of 2.5 minutes.

(a) Speedy Lube guarantees customers that the service will take no longer than 20 minutes. If it does take longer, the customer will receive the service for half-price. What percent of customers receives the service for half price?
(b) If Speedy Lube does not want to give the discount to more than 3% of its customers, how long should it make the guaranteed time limit?

33. Multiple Births The following data represent the relative frequencies of live multiple-delivery births (three or more babies) in 2002 for women 15 to 44 years old.

Age	Relative Frequency
15–19	0.0128
20–24	0.0702
25–29	0.2235
30–34	0.3888
35–39	0.2530
40–44	0.0518

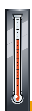
Source: *National Vital Statistics Reports*, Vol. 52, No. 10, December 17, 2003

Suppose the ages of multiple-birth mothers are approximately normally distributed with $\mu = 31.77$ years and standard deviation $\sigma = 5.19$ years.

(a) Compute the proportion of multiple-birth mothers in each class by finding the area under the normal curve.
(b) Compare the proportion to the actual proportions. Are you convinced that the ages of multiple-birth mothers are approximately normally distributed?

34. Weather in Chicago The following frequency distribution represents the daily high temperature in Chicago, November 16 to 30, for the years 1872 to 1999:

(a) Construct a relative frequency distribution.
(b) Draw a relative frequency histogram. Does the distribution of high temperatures appear to be normal?
(c) Compute the mean and standard deviation of high temperature.
(d) Use the information obtained in part (c) to compute the proportion of high temperatures in each class by finding the area under the normal curve.
(e) Are you convinced that high temperatures are approximately normally distributed?



Temperature	Frequency	Temperature	Frequency
5.0–9.9	1	40.0–44.9	375
10.0–14.9	10	45.0–49.9	281
15.0–19.9	15	50.0–54.9	233
20.0–24.9	40	55.0–59.9	160
25.0–29.9	95	60.0–64.9	101
30.0–34.9	217	65.0–69.9	21
35.0–39.9	325	70.0–74.9	1

Source: *Chicago Tribune*. November 27, 2000

Consumer Reports® Sunscreens

More than 1 million skin cancers are expected to be diagnosed in the United States this year, almost half of all cancers diagnosed. The prevalence of skin cancer is attributable in part to a history of unprotected or underprotected sun exposure. Sunscreens have been shown to prevent certain types of lesions associated with skin cancer. They also protect skin against exposure to light that contributes to premature aging. As a result, sunscreen is now in moisturizers, lip balms, shampoos, hair-styling products, insect repellents, and makeup.

Consumer Reports tested 23 sunscreens and two moisturizers, all with a claimed sun-protection factor (SPF) of 15 or higher. SPF is defined as the degree to which a sunscreen protects the skin from UVB, the ultraviolet rays responsible for sunburn. (Some studies have shown that UVB, along with UVA, can increase the risk of skin cancers.) A person with untreated skin who can stay in the sun for 5 minutes before becoming sunburned should be able to stay in the sun for $15 \times 5 = 75$ minutes using a sunscreen rated at SPF15.

To test whether products met their SPF claims for UVB, we used a solar simulator (basically a sun lamp) to expose people to measured amounts of sunlight. First we determined the exposure time (in minutes) that caused each person's untreated skin to turn pink within 24 hours. Then we applied sunscreen to new areas of skin and made the same determination. To avoid potential sources of bias, samples of the sunscreens were applied to randomly assigned sites on the subjects' skin.

To determine the SPF rating of a sunscreen for a particular individual, the exposure time with sunscreen was divided by the exposure time without sunscreen. The following table contains the mean and standard deviation of the SPF measurements for two particular sunscreens.

Product	Mean	Std Dev
A	15.5	1.5
B	14.7	1.2

(a) In designing this experiment, why is it important to obtain the exposure time without sunscreen first and then determine the exposure time with sunscreen for each person?

(b) Why is the random assignment of people and application sites to each treatment (the sunscreen) important?

(c) Calculate the percentage of SPF measurements that you expect to be less than 15, the advertised level of protection for each of the products. (Assume that the SPF ratings are approximately normal.)

(d) Calculate the percentage of SPF measurements that you expect to be greater than 17.5 for product A. Repeat this for product B.

(e) Calculate the percentage of SPF measurements that you expect to fall between 14.5 and 15.5 for product A. Repeat this for product B.

(f) Which product appears to be superior, A or B? Support your conclusion.

Note to Readers: *In many cases, our test protocol and analytical methods are more complicated than described in these examples. The data and discussions have been modified to make the material more appropriate for the audience.*

Technology Step by Step — The Normal Distribution

TI-83/84 Plus

Finding Areas under the Normal Curve

Step 1: From the HOME screen, press 2nd VARS to access the DISTRibution menu.

Step 2: Select `2:normalcdf(`

Step 3: With `normalcdf(` on the HOME screen, type *lowerbound, upperbound*, μ, σ). For example, to find the area to the left of $X = 35$ under the normal curve with $\mu = 40$ and $\sigma = 10$, type

```
Normalcdf(-1E99, 35, 40, 10)
```

and hit ENTER.

Note: When there is no lowerbound, enter −1E99. When there is no upperbound, enter 1E99. The E shown is scientific notation; it is [2nd] [,].

Finding Scores Corresponding to an Area

***Step 1*:** From the HOME screen, press 2^{nd} VARS to access the DISTRibution menu.

***Step 2*:** Select `3:invNorm(`

***Step 3*:** With `invNorm(` on the HOME screen, type "*area left*", μ, σ). For example, to find the score such that the area under the normal curve to the left of the score is 0.68 with $\mu = 40$ and $\sigma = 10$, type

```
InvNorm(0.68, 40, 10)
```

and hit ENTER.

MINITAB

Finding Areas under the Normal Curve

***Step 1*:** Select the **Calc** menu, highlight **Probability Distributions**, and highlight **Normal** . . . .

***Step 2*:** Select **Cumulative Probability**. Enter the mean, μ, and the standard deviation, σ. Select **Input Constant**, and enter the observation. Click OK.

Finding Scores Corresponding to an Area

***Step 1*:** Select the **Calc** menu, highlight **Probability Distributions**, and highlight **Normal** . . . .

***Step 2*:** Select **Inverse Cumulative Probability**. Enter the mean, μ, and the standard deviation, σ. Select **Input Constant**, and enter the area left of the unknown score. Click OK.

Excel

Finding Areas under the Normal Curve

Step 1*:** Select the ***fx button from the tool bar. In **Function Category**:. select "Statistical." In **Function Name**:, select NormDist. Click OK.

***Step 2*:** Enter the specified observation, μ, and σ, and set **cumulative** to True. Click OK.

Finding Scores Corresponding to an Area

Step 1*:** Select the ***fx button from the tool bar. In **Function Category**:, select "Statistical." In **Function Name**:, select NormInv. Click OK.

***Step 2*:** Enter the area left of the unknown score, μ, and σ. Click OK.

7.4 Assessing Normality

Preparing for This Section Before getting started, review the following:

- Shape of a distribution (Section 3.1, pp. 113–116)

Objectives

1 **Draw normal probability plots to assess normality**

Suppose that we obtain a simple random sample from a population whose distribution is unknown. Many of the statistical tests that we perform on small data sets (sample size less than 30) require that the population from which the sample is drawn be normally distributed.

Up to this point, we have said that a random variable X is normally distributed, or at least approximately normal, provided the histogram of the data is symmetric and bell shaped. This method works well for large data sets, but the shape of a histogram drawn from a small sample of observations does not always accurately represent the shape of the population. For this reason, we need additional methods for assessing the normality of a random variable X when we are looking at a small set of sample data.

1 Draw Normal Probability Plots to Assess Normality

In Other Words
Normal probability plots are used to assess normality in small data sets.

A **normal probability plot** is a graph that plots observed data versus *normal scores*. A **normal score** is the expected Z-score of the data value assuming the distribution of the random variable is normal. The expected Z-score of an observed value depends on the number of observations in the data set.

To draw a normal probability plot requires the following steps:

Drawing a Normal Probability Plot

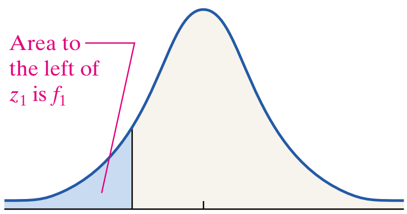
***Step 1*:** Arrange the data in ascending order.

***Step 2*:** Compute $f_i = \dfrac{i - 0.375}{n + 0.25}$,* where i is the index (the position of the data value in the ordered list) and n is the number of observations. The expected proportion of observations less than or equal to the ith data value is f_i.

***Step 3*:** Find the Z-score corresponding to f_i from Table IV.

***Step 4*:** Plot the observed values on the horizontal axis and the corresponding expected Z-scores on the vertical axis.

Figure 42



The idea behind finding the expected Z-score is that, if the data come from a population that is normally distributed, we should be able to predict the area to the left of each data value. The value of f_i represents the expected area to the left of the ith observation when the data come from a population that is normally distributed. For example, f_1 is the expected area to the left of the smallest data value, f_2 is the expected area to the left of the second-smallest data value, and so on. Figure 42 illustrates the idea.

Once we determine each f_i, we find the Z-scores corresponding to f_1, f_2, and so on. The smallest observation in the data set will be the smallest expected Z-score, and the largest observation in the data set will be the largest expected Z-score. Also, because of the symmetry of the normal curve, the expected Z-scores are always paired as positive and negative values.

Normal random variables X and their Z-scores are linearly related $(X = \mu + Z\sigma)$, so a plot of observations of normal variables against their expected normal scores will be linear. We conclude the following:

If sample data are taken from a population that is normally distributed, a normal probability plot of the observed values versus the expected Z-scores will be approximately linear.

Normal probability plots are typically drawn using graphing calculators or statistical software. However, it is worthwhile to go through an example that demonstrates the procedure so that we can better understand the results supplied by software.

EXAMPLE 1

Constructing a Normal Probability Plot

***Problem*:** The data in Table 4 represent the finishing time (in seconds) for six randomly selected races of a greyhound named Barbies Bomber in the $\frac{5}{16}$-mile race at Greyhound Park in Dubuque, Iowa. Is there evidence to support the belief that the variable "finishing time" is normally distributed?

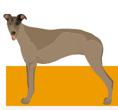


Table 4

31.35	32.52
32.06	31.26
31.91	32.37

Source: Greyhound Park, Dubuque, IA

***Approach*:** We follow Steps 1 through 4 listed above.

*The derivation of this formula is beyond the scope of this text.

Solution

***Step 1*:** The first column in Table 5 represents the index i. The second column represents the observed values in the data set, written in ascending order.

***Step 2*:** The third column in Table 5 represents $f_i = \dfrac{i - 0.375}{n + 0.25}$ for each observation. This is the expected area under the normal curve to the left of the ith observation, assuming normality. For example, $i = 1$ corresponds to the finishing time of 31.26, and

$$f_1 = \frac{1 - 0.375}{6 + 0.25} = 0.1$$

So the area under the normal curve to the left of 31.26 is 0.1 if the sample data come from a population that is normally distributed.

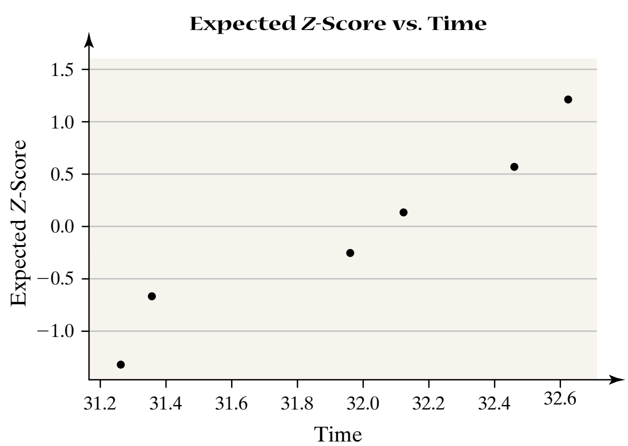
***Step 3*:** We use Table IV to find the Z-scores that correspond to f_i. The expected Z-scores are listed in the fourth column of Table 5. Look in Table IV for the area closest to $f_1 = 0.1$. The expected Z-score is -1.28. Notice that for each negative expected Z-score there is a corresponding positive expected Z-score, as a result of the symmetry of the normal curve.

Table 5

Index, i	Observed Value	f_i	Expected Z-score
1	31.26	$\frac{1 - 0.375}{6 + 0.25} = 0.1$	-1.28
2	31.35	$\frac{2 - 0.375}{6 + 0.25} = 0.26$	-0.64
3	31.91	0.42	-0.20
4	32.06	0.58	0.20
5	32.37	0.74	0.64
6	32.52	0.9	1.28

***Step 4*:** We plot the actual observations on the horizontal axis and the expected Z-scores on the vertical axis. See Figure 43.

Figure 43
Normal Probability Plot



Although the normal probability plot in Figure 43 does show some curvature, it is roughly linear.* We conclude that the finishing time of Barbies Bomber in the $\frac{5}{16}$-mile race is approximately normally distributed.

Typically, normal probability plots are drawn using either a graphing calculator with advanced statistical features or statistical software. Certain software, such as MINITAB, provides bounds that the data must lie within to support the belief that the sample data come from a population that is normally distributed.

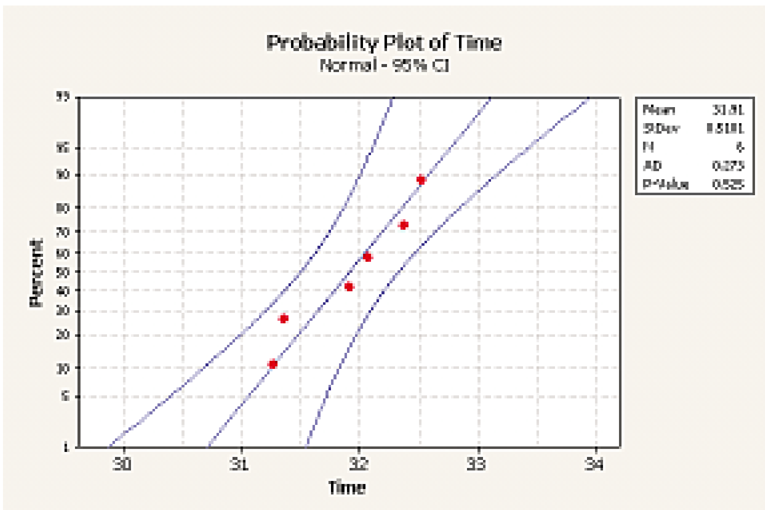
EXAMPLE 2 Assessing Normality Using Technology

***Problem*:** Using MINITAB or some other statistical software, draw a normal probability plot of the data in Table 4 and determine whether there is evidence to support the belief that the sample data come from a population that is normally distributed.

***Approach*:** We will construct a normal probability plot using MINITAB. MINITAB provides curved *bounds* that can be used to assess normality. If the normal probability plot is roughly linear and all the data lie within the bounds provided by the software, we have reason to believe the data come from a population that is approximately normal. The steps for constructing normal probability plots using MINITAB, Excel, or the TI-83/84 Plus graphing calculators can be found on page 361.

***Solution*:** Figure 44 shows the normal probability plot. Notice that MINITAB gives area to the left of the expected Z-score, rather than the Z-score. For example, the area to the left of the expected Z-score of -1.28 is 0.10. MINITAB writes 0.10 as 10 percent.

Figure 44



The normal probability plot is roughly linear, and all the data lie within the bounds provided by MINITAB. We conclude that the sample data could come from a population that is normally distributed.

Throughout the text, we will provide normal probability plots drawn with MINITAB so that assessing normality is straightforward.

*In fact, the correlation between the observed value and expected Z-score is 0.970.

EXAMPLE 3 Assessing Normality

Problem: The data in Table 6 represent the time spent waiting in line (in minutes) for the Demon Roller Coaster for 100 randomly selected riders. Is the random variable "waiting time" normally distributed?

Table 6

7	3	5	107	8	37	16	41	7	25	22	19	1	40	1	29	93
33	76	14	8	9	45	15	81	94	10	115	18	0	18	11	60	34
30	6	21	0	86	6	11	1	1	3	9	79	41	2	9	6	19
4	3	2	7	18	0	93	68	6	94	16	13	24	6	12	121	30
35	39	9	15	53	9	47	5	55	64	51	80	26	24	12	0	
94	18	4	61	38	38	21	61	9	80	18	21	8	14	47	56	

Approach: We will use MINITAB to draw a normal probability plot. If the normal probability plot is roughly linear and the data lie within the bounds provided by MINITAB, conclude that it is reasonable to believe that the sample data come from a population that follows a normal distribution.

Solution: Figure 45 shows a normal probability plot of the data drawn using MINITAB. Since the normal probability plot is not linear, the random variable "wait time" is not normally distributed. Figure 46 shows a histogram of the data in Table 6. The histogram indicates that the data are skewed right.

Figure 45

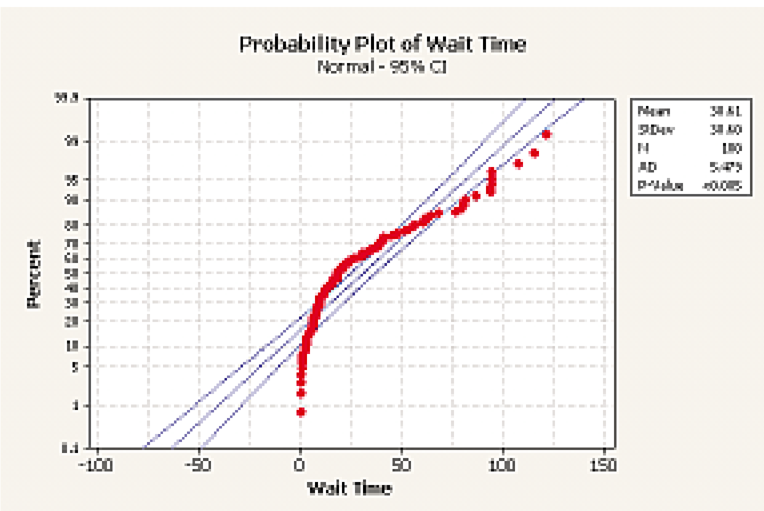
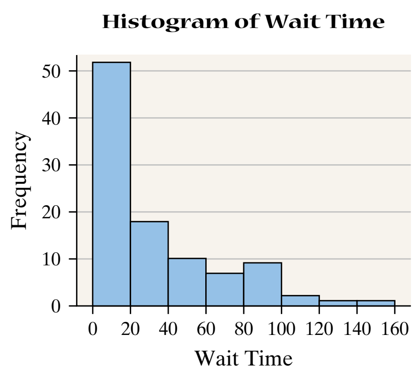


Figure 46



Now Work Problems 3 and 7.

MAKING AN INFORMED DECISION

Join the Club

Suppose that you are interested in starting your own MENSA-type club. To qualify for the club, the potential member must have intelligence that is in the top 20% of all people. The problem that you face is that you do not have a baseline for measuring what qualifies as a top 20% score. To gather these data, you must obtain a random sample of at least 25 volunteers to take an online intelligence test. There are many online intelligence tests, but you need to make sure that the test will supply scored exams. One suggested site is www.queendom.com/tests/iq/classical_iq_r2_access.html.

Once you have obtained your sample of at least 25 test scores, answer the following questions.

(a) What is the mean test score? What is the standard deviation of the test scores?

(b) Do the sample data come from a population that is normally distributed? How do you know this?

(c) Assuming that the sample data come from a population that is normally distributed, determine the test score that would be required to join your club. That is, determine the test score that serves as a cutoff point for the top 20%. You can use this score to determine which potential members may join!

7.4 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

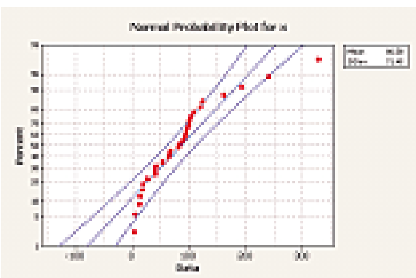
1. Explain why normal probability plots should be linear if the data are normally distributed.

2. What does f_i represent?

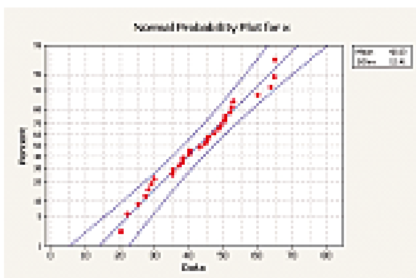
Skill Building

In Problems 3–8, determine whether the normal probability plot indicates that the sample data could have come from a population that is normally distributed.

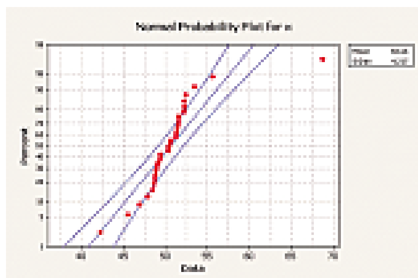
3. NW



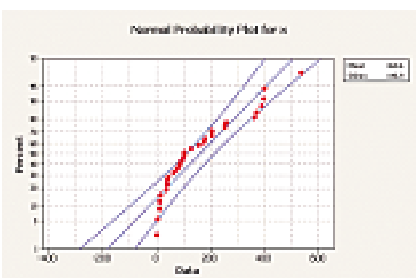
4.



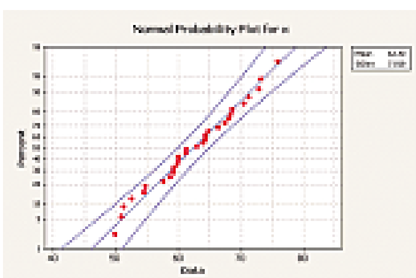
5.



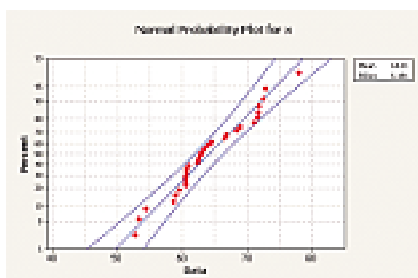
6.



7. NW



8.

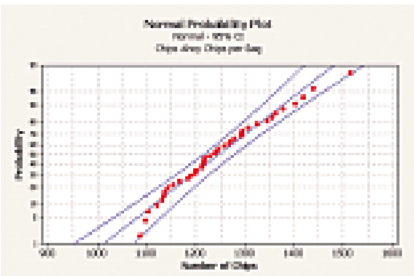


Applying the Concepts

9. Chips per Bag In a 1998 advertising campaign, Nabisco claimed that every 18-ounce bag of Chips Ahoy! cookies contained at least 1000 chocolate chips. Brad Warner and Jim Rutledge (*Chance*, Vol. 12, No. 1, 1999) tried to verify the claim. The following data represent the number of chips in an 18-ounce bag of Chips Ahoy! based on their study.

1087	1098	1103	1121	1132
1185	1191	1199	1200	1213
1239	1244	1247	1258	1269
1307	1325	1345	1356	1363
1135	1137	1143	1154	1166
1214	1215	1219	1219	1228
1270	1279	1293	1294	1295
1377	1402	1419	1440	1514

(a) Use the following normal probability plot to determine if the data could have come from a normal distribution.



(b) Determine the mean and standard deviation of the sample data.

(c) Using the sample mean and sample standard deviation obtained in part (b) as estimates for the population mean and population standard deviation, respectively, draw a graph of a normal model for the distribution of chips in a bag a Chips Ahoy!.

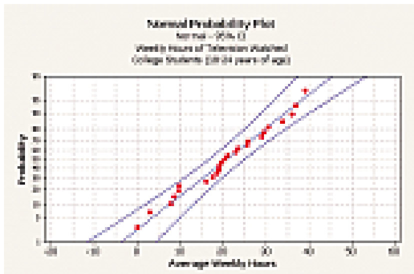
(d) Using the normal model from part (c), find the probability that an 18-ounce bag of Chips Ahoy! selected at random contains at least 1000 chips.

(e) Using the normal model from part (c), determine the proportion of 18-ounce bags of Chips Ahoy! that contains between 1200 and 1400 chips.

10. Hours of TV A random sample of college students aged 18 to 24 years was obtained, and the number of hours of television watched last week was recorded.

36.1	30.5	2.9	17.5	21.0
23.5	25.6	16.0	28.9	29.6
7.8	20.4	33.8	36.8	0.0
9.9	25.8	19.5	19.1	18.5
22.9	9.7	39.2	19.0	8.6

(a) Use the following normal probability plot to determine if the data could have come from a normal distribution.



(b) Determine the mean and standard deviation of the sample data.

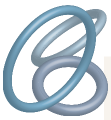
(c) Using the sample mean and sample standard deviation obtained in part (b) as estimates for the population mean and population standard deviation, respectively, draw a graph of a normal model for the distribution of weekly hours of television watched.

(d) Using the normal model from part (c), find the probability that a college student aged 18 to 24 years, selected at random, watches between 20 and 35 hours of television each week.

(e) Using the normal model from part (c), determine the proportion of college students aged 18 to 24 years who watch more than 40 hours of television per week.

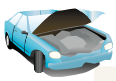
In Problems 11–14, use a normal probability plot to assess whether the sample data could have come from a population that is normally distributed.

11. O-Ring Thickness A random sample of O-rings was obtained and the wall thickness (in inches) of each was recorded.



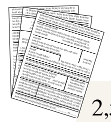
0.276	0.274	0.275	0.274	0.277
0.273	0.276	0.276	0.279	0.274
0.273	0.277	0.275	0.277	0.277
0.276	0.277	0.278	0.275	0.276

12. Customer Service A random sample of weekly work logs at an automobile repair station was obtained and the average number of customers per day was recorded.



26	24	22	25	23
24	25	23	25	22
21	26	24	23	24
25	24	25	24	25
26	21	22	24	24

13. School Loans A random sample of 20 undergraduate students receiving student loans was obtained, and the amount of their loans for the 2004–2005 school year was recorded.



2,500	1,000	2,000	14,000	1,800
3,800	10,100	2,200	900	1,600
500	2,200	6,200	9,100	2,800
2,500	1,400	13,200	750	12,000

14. Memphis Snowfall A random sample of 25 years between 1890 and 2005 was obtained, and the amount of snowfall, in inches, for Memphis was recorded.

(*Source*: National Oceanic and Atmospheric Administration)

24.0	7.9	1.5	0.0	0.3
0.4	8.1	4.3	0.0	0.5
3.6	2.9	0.4	2.6	0.1
16.6	1.4	23.8	25.1	1.6
12.2	14.8	0.4	3.7	4.2

Technology Step by Step — Normal Probability Plots

TI-83/84 Plus

***Step 1*:** Enter the raw data into L1.

***Step 2*:** Press 2nd Y = to access STAT PLOTS.

***Step 3*:** Select `1:Plot1`.

***Step 4*:** Turn `Plot1` ON by highlighting ON and pressing ENTER. Press the down-arrow key. Highlight the *normal probability plot* icon. It is the icon in the lower-right corner under Type:. Press ENTER to select this plot type. The Data List should be set at L1. The data axis should be the x-axis.

***Step 5*:** Press ZOOM, and select `9:ZoomStat`.

MINITAB

***Step 1*:** Enter the raw data into C1.

***Step 2*:** Select the **Graph** menu. Highlight **Probability Plot** ....

***Step 3*:** In the Variables cell, enter the column that contains the raw data. Make sure Distribution is set to Normal. Click OK.

Excel

***Step 1*:** Load the PHStat Add-in.

***Step 2*:** Enter the raw data into column A.

***Step 3*:** Select the **PHStat** menu. Highlight **Probability Distributions**, then highlight **Normal Probability Plot** ....

***Step 4*:** With the cursor in the "Variable Cell Range:" cell, highlight the raw data. Enter a graph title, if desired. Click OK.

7.5 The Normal Approximation to the Binomial Probability Distribution

Preparing for This Section Before getting started, review the following:

- Binomial probability distribution (Section 6.2, pp. 298–309)

Objective 1 **Approximate binomial probabilities using the normal distribution**

1 Approximate Binomial Probabilities Using the Normal Distribution

In Section 6.2, we discussed the binomial probability distribution. A probability experiment is said to be a binomial experiment if the following conditions are met.

Criteria for a Binomial Probability Experiment

A probability experiment is said to be a binomial experiment if all the following are true:

1. The experiment is performed n independent times. Each repetition of the experiment is called a **trial**. Independence means that the outcome of one trial will not affect the outcome of the other trials.
2. For each trial, there are two mutually exclusive outcomes—success or failure.
3. The probability of success, p, is the same for each trial of the experiment.

The binomial probability formula can be used to compute probabilities of events in a binomial experiment. When there is a large number of trials of a binomial experiment, the binomial probability formula can be difficult to use. For example, suppose there are 500 trials of a binomial experiment and we wish to compute the probability of 400 or more successes. Using the binomial probability formula requires that we compute the following probabilities:

$$P(X \geq 400) = P(400) + P(401) + \cdots + P(500)$$

This would be time consuming to compute by hand! Fortunately, we have other means for approximating binomial probabilities, provided that certain conditions are met.

Recall, as the number of trials, n, in a binomial experiment increases, the probability histogram becomes more nearly symmetric and bell shaped (see page 308). We restate the conclusion here.

As the number of trials n in a binomial experiment increases, the probability distribution of the random variable X becomes more nearly symmetric and bell shaped. As a rule of thumb, if $np(1 - p) \geq 10$, the probability distribution will be approximately symmetric and bell shaped.

Because of this result, we might be inclined to think that binomial probabilities can be approximated by the area under the normal curve, provided that $np(1 - p) \geq 10$. This intuition is correct.

Historical Note

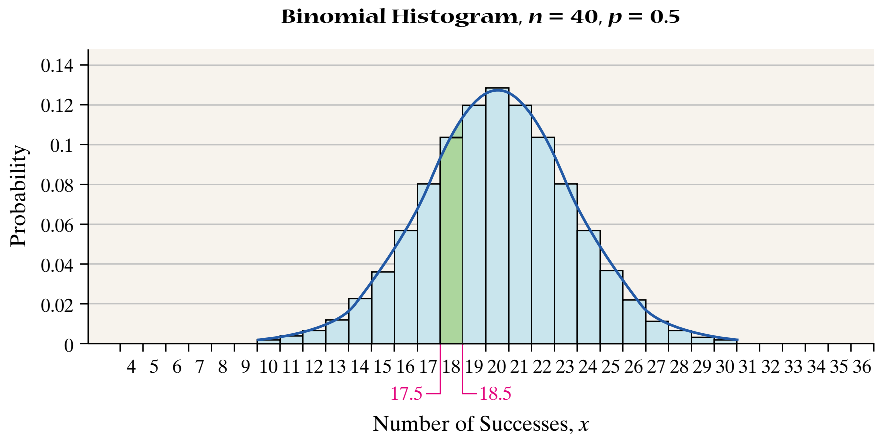
The normal approximation to the binomial was discovered by Abraham de Moivre in 1733. With the advance of computing technology, its importance has been diminished.

The Normal Approximation to the Binomial Probability Distribution

If $np(1 - p) \geq 10$, the binomial random variable X is approximately normally distributed with mean $\mu_X = np$ and standard deviation $\sigma_X = \sqrt{np(1 - p)}$.

Figure 47 shows a probability histogram for the binomial random variable X with $n = 40$ and $p = 0.5$ and a normal curve with $\mu_X = np = 40(0.5) = 20$ and standard deviation $\sigma_X = \sqrt{np(1 - p)} = \sqrt{40(0.5)(0.5)} = \sqrt{10}$. Notice that $np\,(1 - p) = 40(0.5)(1 - 0.5) = 10$.

Figure 47

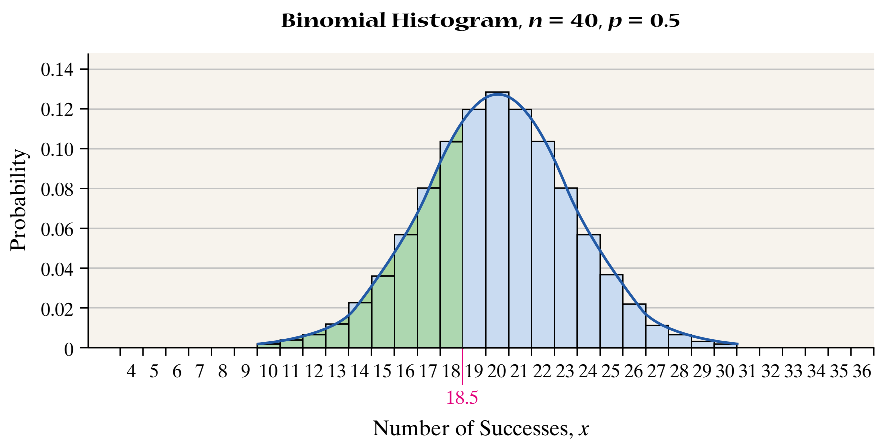


CAUTION
Don't forget about the correction for continuity. It is needed because we are using a continuous density function to approximate the probability of a discrete random variable.

We know from Section 6.2 that the area of the rectangle corresponding to $X = 18$ represents $P(18)$. The width of each rectangle is 1, so the rectangle extends from $X = 17.5$ to $X = 18.5$. The area under the normal curve from $X = 17.5$ to $X = 18.5$ is approximately equal to the area of the rectangle corresponding to $X = 18$. Therefore, the area under the normal curve between $X = 17.5$ and $X = 18.5$ is approximately equal to $P(18)$, where X is a binomial random variable with $n = 40$ and $p = 0.5$. We add and subtract 0.5 from $X = 18$ as a **correction for continuity**, because we are using a continuous density function to approximate a discrete probability.

Suppose we want to approximate $P(X \leq 18)$. Figure 48 illustrates the situation.

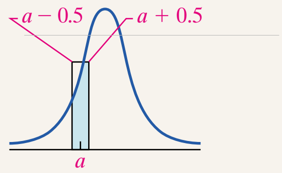
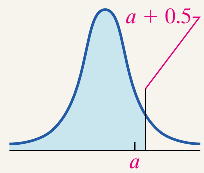
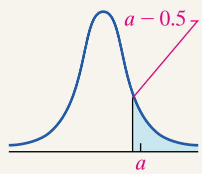
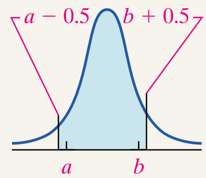
Figure 48



To approximate $P(X \leq 18)$, we compute the area under the normal curve for $X < 18.5$. Do you see why?

If we want to approximate $P(X \geq 18)$, we compute $P(X \geq 17.5)$. Do you see why? Table 7 summarizes how to use the correction for continuity.

Table 7

Exact Probability Using Binomial	Approximate Probability Using Normal	Graphical Depiction
$P(X = a)$	$P(a - 0.5 < X < a + 0.5)$	
$P(X \leq a)$	$P(X < a + 0.5)$	
$P(X \geq a)$	$P(X > a - 0.5)$	
$P(a \leq X \leq b)$	$P(a - 0.5 < X < b + 0.5)$	

A question remains, however. What do we do if the probability is of the form $P(X > a)$, $P(X < a)$, or $P(a < X < b)$? The solution is to rewrite the inequality in a form with $\leq$ or $\geq$. For example, $P(X > 4) = P(X \geq 5)$ and $P(X < 4) = P(X \leq 3)$ for binomial random variables, because the values of the random variables must be whole numbers.

EXAMPLE 1 The Normal Approximation to a Binomial Random Variable

Problem: According to the *Information Please Almanac*, 6% of the human population has blood type O-negative. What is the probability that, in a simple random sample of 500, fewer than 25 have blood type O-negative?

Approach

Step 1: We verify that this is a binomial experiment.

Step 2: Computing the probability by hand would be very tedious. Verify $np(1 - p) \geq 10$. Then we will know that the condition for using the normal distribution to approximate the binomial distribution is met.

Step 3: Approximate $P(X < 25) = P(X \leq 24)$ by using the normal approximation to the binomial distribution.

Solution

Step 1: There are 500 independent trials with each trial having a probability of success equal to 0.06. This is a binomial experiment.

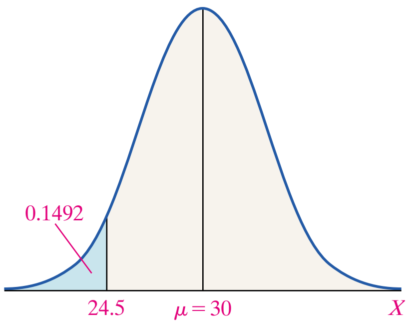
Step 2: We verify $np(1 - p) \geq 10$.

$$np(1 - p) = 500(0.06)(0.94) = 28.2 \geq 10$$

We can use the normal distribution to approximate the binomial distribution.

***Step 3*:** We wish to know the probability that fewer than 25 people in the sample have blood type O-negative; that is, we wish to know $P(X < 25) = P(X \le 24)$. This is approximately equal to the area under the normal curve to the left of $X = 24.5$, with $\mu_X = np = 500(0.06) = 30$ and $\sigma_X = \sqrt{np(1-p)} = \sqrt{500(0.06)(1-0.06)} = \sqrt{28.2} \approx 5.31$. See Figure 49. We convert $X = 24.5$ to a Z-score.

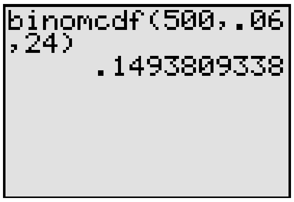
Figure 49



$$Z = \frac{24.5 - 30}{5.31} = -1.04$$

From Table IV, we find the area to the left of $Z = -1.04$ is 0.1492. Therefore, the approximate probability that fewer than 25 people will have blood type O-negative is $0.1492 = 14.92\%$.

Figure 50



Using the *binomcdf(* command on a TI-84 Plus graphing calculator, we find that the exact probability is 0.1494. See Figure 50. The approximate result is close indeed!

Now Work Problem 21.

EXAMPLE 2 A Normal Approximation to the Binomial

***Problem*:** According to the Federal Communications Commission, 70% of all U.S. households have cable television. Erica conducts a random sample of 1000 households in DuPage County and finds that 734 of them have cable.

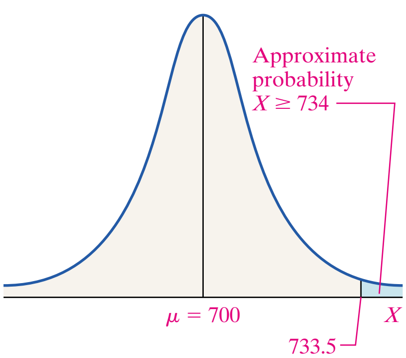
(a) Assuming that 70% of households have cable, what is the probability of obtaining a random sample of at least 734 households with cable from a sample of size 1000?

(b) Does the result from part (a) contradict the FCC information? Explain.

***Approach*:** This is a binomial experiment with $n = 1000$ and $p = 0.70$. Erica needs to determine the probability of obtaining a random sample of at least 734 households with cable from a sample of size 1000, assuming 70% of households have cable. Computing this using the binomial probability formula would be difficult, so Erica will compute the probability using the normal approximation to the binomial, since $np(1-p) = 1000(0.70)(0.30) = 210 \ge 10$. We approximate $P(X \ge 734)$ by computing the area under the standard normal curve to the right of $X = 733.5$ with $\mu_X = np = 1000(0.70) = 700$ and $\sigma_X = \sqrt{np(1-p)} = \sqrt{1000(0.70)(1-0.70)} = \sqrt{210} \approx 14.491$.

Solution

Figure 51



(a) Figure 51 shows the area we wish to compute. We convert $X = 733.5$ to a Z-score.

$$Z = \frac{733.5 - 700}{14.491} = 2.31$$

The area under the standard normal curve to the right of $Z = 2.31$ is $1 - 0.9896 = 0.0104$. There is a 1.04% probability of obtaining 734 or more households with cable from a sample of 1000 households, assuming that the percentage of households with cable is 70%.

(b) Yes. The result from part (a) means that about 1 sample in every 100 samples will have 734 or more households with cable if the true proportion is 0.7. Erica is not inclined to believe that her sample is one of the 1 in 100. She would rather believe that the proportion of households in DuPage County with cable is higher than 0.70.

Now Work Problem 27.

7.5 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. List the conditions required for a binomial experiment.

2. Under what circumstances can the normal distribution be used to approximate binomial probabilities?

3. Why must we use a correction for continuity when using the normal distribution to approximate binomial probabilities?

4. *True or False*: Suppose X is a binomial random variable. To approximate $P(3 \leq X < 7)$ using the normal probability distribution, we compute $P(3.5 \leq X < 7.5)$.

Skill Building

In Problems 5–14, a discrete random variable is given. Assume the probability of the random variable will be approximated using the normal distribution. Describe the area under the normal curve that will be computed. For example, if we wish to compute the probability of finding at least five defective items in a shipment, we would approximate the probability by computing the area under the normal curve to the right of $X = 4.5$.

5. The probability that at least 40 households have a gas stove

6. The probability of no more than 20 people who want to see *Roe v. Wade* overturned

7. The probability that exactly eight defective parts are in the shipment

8. The probability that exactly 12 students pass the course

9. The probability that the number of people with blood type O-negative is between 18 and 24, inclusive

10. The probability that the number of tornadoes that occur in the month of May is between 30 and 40, inclusive

11. The probability that more than 20 people want to see the marriage tax penalty abolished

12. The probability that fewer than 40 households have a pet

13. The probability that more than 500 adult Americans support a bill proposing to extend daylight savings time

14. The probability that fewer than 35 people support the privatization of Social Security

In Problems 15–20, compute P(x) using the binomial probability formula. Then determine whether the normal distribution can be used as an approximation for the binomial distribution. If so, approximate P(x) and compare the result to the exact probability.

15. $n = 60, p = 0.4, X = 20$

16. $n = 80, p = 0.15, X = 18$

17. $n = 40, p = 0.25, X = 30$

18. $n = 100, p = 0.05, X = 50$

19. $n = 75, p = 0.75, X = 60$

20. $n = 85, p = 0.8, X = 70$

Applying the Concepts

21. On-Time Flights According to American Airlines, Flight
NW 215 from Orlando to Los Angeles is on time 90% of the time. Suppose 150 flights are randomly selected. Use the normal approximation to the binomial to

(a) approximate the probability that exactly 130 flights are on time.

(b) approximate the probability that at least 130 flights are on time.

(c) approximate the probability that fewer than 125 flights are on time.

(d) approximate the probability that between 125 and 135 flights, inclusive, are on time.

22. Smokers According to *Information Please Almanac*, 80% of adult smokers started smoking before they were 18 years old. Suppose 100 smokers 18 years old or older are randomly selected. Use the normal approximation to the binomial to

(a) approximate the probability that exactly 80 of them started smoking before they were 18 years old.

(b) approximate the probability that at least 80 of them started smoking before they were 18 years old.

(c) approximate the probability that fewer than 70 of them started smoking before they were 18 years old.

(d) approximate the probability that between 70 and 90 of them, inclusive, started smoking before they were 18 years old.

23. Migraine Sufferers In clinical trials of a medication whose purpose is to reduce the pain associated with migraine headaches, 2% of the patients in the study experienced weight gain as a side effect. Suppose a random sample of 600 users of this medication is obtained. Use the normal approximation to the binomial to

(a) approximate the probability that exactly 20 will experience weight gain as a side effect.

(b) approximate the probability that 20 or fewer will experience weight gain as a side effect.

(c) approximate the probability that 22 or more patients will experience weight gain as a side effect.

(d) approximate the probability that between 20 and 30 patients, inclusive, will experience weight gain as a side effect.

24. High-Speed Internet According to a report by the Commerce Department in the fall of 2004, 20% of U.S. households had some type of high-speed Internet connection. Suppose 80 U.S. households are selected at random. Use the normal approximation to the binomial to

(a) approximate the probability that exactly 15 households have high-speed Internet access.

(b) approximate the probability that at least 20 households have high-speed Internet access.

(c) approximate the probability that fewer than 10 households have high-speed Internet access.
(d) approximate the probability that between 12 and 18 households, inclusive, have high-speed Internet access.

25. Allergy Sufferers Clarinex-D is a medication whose purpose is to reduce the symptoms associated with a variety of allergies. In clinical trials of Clarinex-D, 5% of the patients in the study experienced insomnia as a side effect. Suppose a random sample of 400 Clarinex-D users is obtained. Use the normal approximation to the binomial to
(a) approximate the probability that exactly 20 patients experienced insomnia as a side effect.
(b) approximate the probability that 15 or fewer patients experienced insomnia as a side effect.
(c) approximate the probability that 30 or more patients experienced insomnia as a side effect.
(d) approximate the probability that between 10 and 32 patients, inclusive, experienced insomnia as a side effect.

26. Murder by Firearms According to the *Uniform Crime Report, 2003*, 66.9% of murders are committed with a firearm. Suppose that 50 murders are randomly selected. Use the normal approximation to the binomial to
(a) approximate the probability that exactly 40 murders are committed using a firearm.
(b) approximate the probability that at least 35 murders are committed using a firearm.
(c) approximate the probability that fewer than 25 murders are committed using a firearm.
(d) approximate the probability that between 30 and 35 murders, inclusive, are committed using a firearm.

27. Males Living at Home NW According to the *Current Population Survey* (Internet release date: September 15, 2004), 55% of males between the ages of 18 and 24 years lived at home in 2003. (Unmarried college students living in a dorm are counted as living at home.) Suppose that a survey is administered at a community college to 200 randomly selected male students between the ages of 18 and 24 years and that 130 of them respond that they live at home.
(a) Approximate the probability that such a survey will result in at least 130 of the respondents living at home under the assumption that the true percentage is 55%.
(b) Does the result from part (a) contradict the results of the *Current Population Survey*? Explain.

28. Females Living at Home According to the *Current Population Survey* (Internet release date: September 15, 2004), 46% of females between the ages of 18 and 24 years lived at home in 2003. (Unmarried college students living in a dorm are counted as living at home.) Suppose that a survey is administered at a community college to 200 randomly selected female students between the ages of 18 and 24 years and that 110 of them respond that they live at home.
(a) Approximate the probability that such a survey will result in at least 110 of the respondents living at home under the assumption that the true percentage is 46%.
(b) Does the result from part (a) contradict the results of the *Current Population Survey*? Explain.

29. Boys Are Preferred In a Gallup poll conducted December 2–4, 2000, 42% of survey respondents said that, if they only had one child, they would prefer the child to be a boy. Suppose you conduct a survey of 150 randomly selected students on your campus and find that 80 of them would prefer a boy.
(a) Approximate the probability that, in a random sample of 150 students, at least 80 would prefer a boy, assuming the true percentage is 42%.
(b) Does this result contradict the Gallup poll? Explain.

30. Liars According to a *USA Today* "Snapshot," 3% of Americans surveyed lie frequently. Suppose you conduct a survey of 500 college students and find that 20 of them lie frequently.
(a) Compute the probability that, in a random sample of 500 college students, at least 20 lie frequently, assuming the true percentage is 3%.
(b) Does this result contradict the *USA Today* "Snapshot"? Explain.

CHAPTER 7 Review

Summary

In this chapter, we introduced continuous random variables and the normal probability density function. A continuous random variable is said to be approximately normally distributed if a histogram of its values is symmetric and bell shaped. In addition, we can draw normal probability plots that are based on expected Z-scores. If these normal probability plots are approximately linear, we say the distribution of the random variable is approximately normal. The area under the normal density function can be used to find proportions, probabilities or percentiles for normal random variables. Also, we can find the value of a normal random variable that corresponds to a specific proportion, probability, or percentile.

If X is a binomial random variable with $np(1-p) \geq 10$, we can use the area under the normal curve to approximate the probability of a binomial random variable. The parameters of the normal curve are $\mu_X = np$ and $\sigma_X = \sqrt{np(1-p)}$, where n is the number of trials of the binomial experiment and p is the probability of success.

Formulas

Standardizing a Normal Random Variable

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

Finding the Score

$$X = \mu + Z\sigma$$

Vocabulary

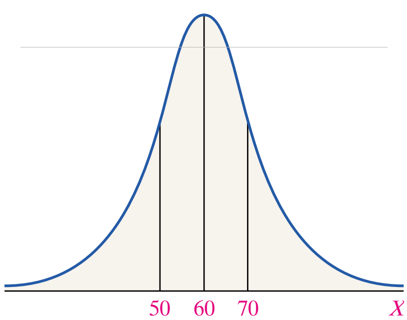
Uniform probability distribution (p. 319)
Probability density function (p. 320)
Normal curve (p. 321)
Normal probability distribution (p. 321)
Inflection points (p. 321)
Normal probability density function (p. 324)
Standard normal distribution (p. 326)
Normal probability plot (p. 355)
Normal score (p. 355)
Trial (p. 362)
Correction for continuity (p. 363)

Objectives

Section	You should be able to . . .	Examples	Review Exercises
7.1	1 Understand the uniform probability distribution (p. 319)	1 and 2	37
	2 Graph a normal curve (p. 321)	Page 321	19–22
	3 State the properties of the normal curve (p. 322)	Page 322	38
	4 Understand the role of area in the normal density function (p. 323)	3 and 4	1, 2
	5 Understand the relation between a normal random variable and a standard normal random variable (p. 325)	5	3, 4
7.2	1 Find the area under the standard normal curve (p. 332)	1 through 4	5–8
	2 Find the Z-scores for a given area (p. 336)	5 through 9	13–18
	3 Interpret the area under standard normal curve as a probability (p. 340)	10	9–12
7.3	1 Find and interpret the area under a normal curve (p. 345)	1 through 3	19–22, 23(a)–(c), 24(a)–(c), 25(a)–(c), 26(a)–(d), 27, 28
	2 Find the value of a normal random variable (p. 348)	4 through 6	23(d), 24(d), (e), 25(d), (e), 26(e), (f)
7.4	1 Draw normal probability plots to assess normality (p. 355)	1 through 3	31–34
7.5	1 Approximate binomial probabilities using the normal distribution (p. 362)	1 and 2	29, 30

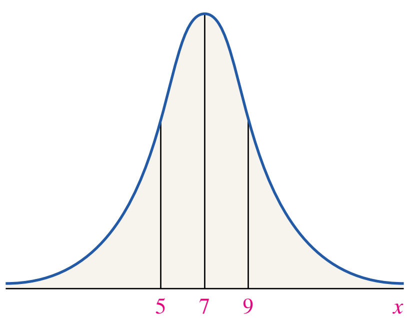
Review Exercises

1. Use the figure to answer the questions that follow:



(a) What is μ?
(b) What is σ?
(c) Suppose the area under the normal curve to the right of $X = 75$ is 0.0668. Provide two interpretations for this area.
(d) Suppose the area under the normal curve between $X = 50$ and $X = 75$ is 0.7745. Provide two interpretations for this area.

2. Use the figure to answer the questions that follow:



(a) What is μ?
(b) What is σ?
(c) Suppose the area under the normal curve to the left of $X = 10$ is 0.9332. Provide two interpretations for this area.
(d) Suppose the area under the normal curve between $X = 5$ and $X = 8$ is 0.5328. Provide two interpretations for this area.

3. A random variable X is approximately normally distributed with $\mu = 20$ and $\sigma = 4$.
   (a) Compute $Z_1 = \dfrac{X_1 - \mu}{\sigma}$ for $X_1 = 18$.
   (b) Compute $Z_2 = \dfrac{X_2 - \mu}{\sigma}$ for $X_2 = 21$.
   (c) The area under the normal curve between $X_1 = 18$ and $X_2 = 21$ is 0.2912. What is the area between Z_1 and Z_2?

4. A random variable X is approximately normally distributed with $\mu = 50$ and $\sigma = 8$.
   (a) Compute $Z_1 = \dfrac{X_1 - \mu}{\sigma}$ for $X_1 = 48$.
   (b) Compute $Z_2 = \dfrac{X_2 - \mu}{\sigma}$ for $X_2 = 60$.
   (c) The area under the normal curve between $X_1 = 48$ and $X_2 = 60$ is 0.4931. What is the area between Z_1 and Z_2?

In Problems 5–8, draw a standard normal curve and shade the area indicated. Then use Table IV to find the area under the normal curve.

5. The area to the left of $Z = -1.04$
6. The area to the right of $Z = 2.04$
7. The area between $Z = -0.34$ and $Z = 1.03$
8. The area between $Z = 1.93$ and $Z = 3.93$

In Problems 9–12, find the indicated probability of the standard normal random variable Z.

9. $P(Z < 1.19)$
10. $P(Z \geq 1.61)$
11. $P(-1.21 < Z \leq 2.28)$
12. $P(0.21 < Z < 1.69)$
13. Find the Z-score such that the area to the left of the Z-score is 0.84.
14. Find the Z-score such that the area to the right of the Z-score is 0.483.
15. Find the Z-scores that separate the middle 92% of the data from the area in the tails of the standard normal distribution.
16. Find the Z-scores that separate the middle 88% of the data from the area in the tails of the standard normal distribution.
17. Find the value of $z_{0.20}$.
18. Find the value of $z_{0.04}$.

In Problems 19–22, draw the normal curve with the parameters indicated. Then find the probability of the random variable X. Shade the area that represents the probability.

19. $\mu = 50, \sigma = 6, P(X > 55)$
20. $\mu = 30, \sigma = 5, P(X \leq 23)$
21. $\mu = 70, \sigma = 10, P(65 < X < 85)$
22. $\mu = 20, \sigma = 3, P(22 \leq X \leq 27)$

23. **Tire Wear** Suppose Dunlop Tire manufactures tires having the property that the mileage the tire lasts approximately follows a normal distribution with mean 70,000 miles and standard deviation 4400 miles.
   (a) What percent of the tires will last at least 75,000 miles?
   (b) Suppose Dunlop warrants the tires for 60,000 miles. What percent of the tires will last 60,000 miles or less?
   (c) What is the probability that a randomly selected Dunlop tire lasts between 65,000 and 80,000 miles?
   (d) Suppose that Dunlop wants to warrant no more than 2% of its tires. What mileage should the company advertise as its warranty mileage?

24. **Talk Time on a Cell Phone** Suppose the talk time in digital mode on a Motorola Timeport P8160 is approximately normally distributed with mean 324 minutes and standard deviation 24 minutes.
   (a) What proportion of the time will a fully charged battery last at least 300 minutes?
   (b) What proportion of the time will a fully charged battery last less than 340 minutes?
   (c) Suppose you charge the battery fully. What is the probability it will last between 310 and 350 minutes?
   (d) Determine the talk time that is in the top 20%.
   (e) Determine the talk time that makes up the middle 90% of talk time.

25. **Serum Cholesterol** As reported by the U.S. National Center for Health Statistics, the mean serum cholesterol of females 16 to 19 years old is $\mu = 171$. If serum cholesterol is approximately normally distributed with $\sigma = 39.8$, answer the following:
   (a) Determine the proportion of 16- to 19-year-old females with a serum cholesterol above 180.
   (b) Determine the proportion of 16- to 19-year-old females with a serum cholesterol between 150 and 200.
   (c) Suppose a 16- to 19-year-old female is randomly selected. Determine the probability her serum cholesterol is below 140.
   (d) Determine the serum cholesterol that divides the bottom 10% from the top 90% of all serum cholesterol levels of 16- to 19-year-old females.
   (e) According to the National Center for Health Statistics, the 25th percentile of serum cholesterol for 16- to 19-year-old females is 145. Is the 25th percentile on the normal curve close to the reported value of 145?

26. **Wechsler Intelligence Scale** The Wechsler Intelligence Scale for Children is approximately normally distributed with mean 100 and standard deviation 15.
   (a) What proportion of test takers will score above 125?
   (b) What proportion of test takers will score below 90?
   (c) What proportion of test takers will score between 110 and 140?
   (d) If a child is randomly selected, what is the probability that she scores above 150?
   (e) What intelligence score will place a child in the top 5% of all children?
   (f) If normal intelligence is defined as scoring in the middle 95% of all test takers, figure out the scores that differentiate normal intelligence from abnormal intelligence.

27. Major League Baseballs According to major league baseball rules, the ball must weigh between 5 and 5.25 ounces. (*Source*: www.baseball-almanac.com) Suppose a factory produces baseballs whose weights are approximately normally distributed with mean 5.11 ounces and standard deviation 0.062 ounces.

(a) What proportion of the baseballs produced by this factory are too heavy for use by major league baseball?
(b) What proportion of the baseballs produced by this factory are too light for use by the major league baseball?
(c) What proportion of the baseballs produced by this factory can be used by major league baseball?

28. Halogen Light Bulbs Feit Electric manufactures a Crystal Clear Halogen 60-watt light bulb that has an average life of 3000 hours. Suppose the lifetimes of the light bulbs are approximately normally distributed with standard deviation 183 hours.

(a) What proportion of the light bulbs will last more than 3300 hours?
(b) What proportion of the light bulbs will last less than 2500 hours?
(c) What is the probability that a randomly selected Feit Crystal Clear Halogen 60-watt light bulb lasts between 2900 and 3100 hours?
(d) What is the probability that a randomly selected Feit Crystal Clear Halogen 60-watt light bulb lasts less than 2600 hours?
(e) What is the percentile rank of a Feit Crystal Clear Halogen 60-watt light bulb that lasts 3350 hours?
(f) Would it be unusual for a Feit Crystal Clear Halogen 60-watt light bulb to last longer than 3400 hours?

29. High Cholesterol According to the National Center for Health Statistics, 8% of 20- to 34-year-old females have high serum cholesterol. Suppose you conduct a random sample of two hundred 20- to 34-year-old females.

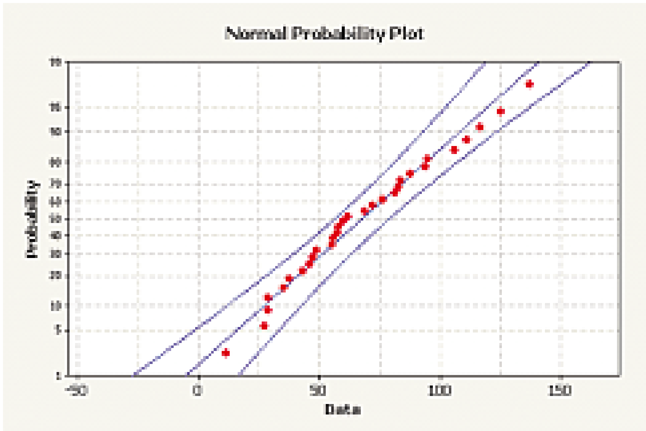
(a) Verify that the conditions for using the normal distribution to approximate the binomial distribution are met.
(b) Approximate the probability that exactly 15 have high serum cholesterol. Interpret this result.
(c) Approximate the probability that more than 20 have high serum cholesterol. Interpret this result.
(d) Approximate the probability that at least 15 have high serum cholesterol. Interpret this result.
(e) Approximate the probability that fewer than 25 have high serum cholesterol. Interpret this result.
(f) Approximate the probability that between 15 and 25, inclusive, have high serum cholesterol. Interpret this result.

30. America Reads According to a Gallup poll conducted September 10–14, 1999, 56% of Americans 18 years old or older stated they had read at least six books (fiction and nonfiction) within the past year. Suppose you conduct a random sample of 250 Americans 18 years old or older.

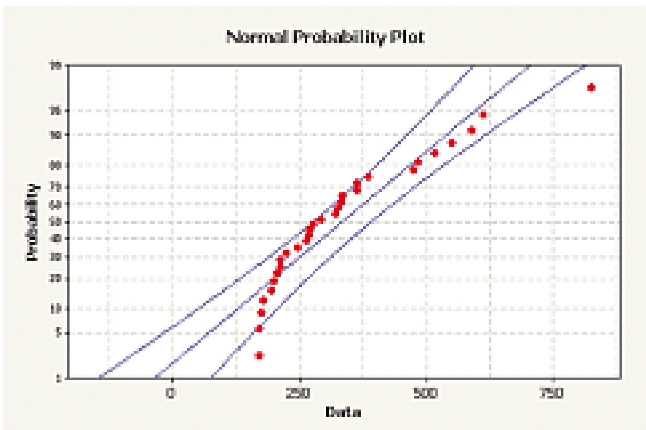
(a) Verify that the conditions for using the normal distribution to approximate the binomial distribution are met.
(b) Approximate the probability that exactly 125 read at least six books within the past year. Interpret this result.
(c) Approximate the probability that fewer than 120 read at least six books within the past year. Interpret this result.
(d) Approximate the probability that at least 140 read at least six books within the past year. Interpret this result.
(e) Approximate the probability that between 100 and 120, inclusive, read at least six books within the past year. Interpret this result.

In Problems 31 and 32, a normal probability plot of a simple random sample of data from a population whose distribution is unknown was obtained. Given the normal probability plot, is there reason to believe the population is normally distributed?

31.



32.



In Problems 33 and 34, assess the normality of the sample data.

33. Density of Earth In 1798, Henry Cavendish obtained 27 measurements of the density of Earth, using a torsion balance. The following data represent his estimates, represented as a multiple of the density of water.

5.50	5.57	5.42	5.61	5.53
5.47	4.88	5.62	5.63	4.07
5.29	5.34	5.26	5.44	5.55
5.34	5.30	5.36	5.79	5.29
5.10	5.86	5.58	5.27	5.85
5.65	5.39			

Source: Stigler, S. M. "Do robust estimators work with real data?" *Annals of Statistics,* 5 (1977), 1055–1078.

34. Life Expectancy The following data represent the life expectancy at birth in 2005 in a random sample of 20 countries.

75.84	77.11	75.91	80.39	78.92
74.23	78.62	77.79	77.83	80.10
79.95	76.58	76.84	79.14	77.62
79.21	78.35	77.71	80.40	78.19

Source: U.S. Census Bureau, International Database

35. Hybrid SUV As the first hybrid sports utility vehicle (SUV) with gasoline mileage certified by the Environmental Protection Agency, the Ford Escape HEV is the most fuel-efficient SUV in 2005. The mean mileage for the automatic four-wheel-drive Ford Escape HEV is 29 miles per gallon on the highway. (*Source*: www.fueleconomy.gov) Suppose the gasoline mileages of these SUVs are normally distributed with standard deviation 2 miles per gallon.

(a) What proportion of automatic four-wheel-drive 2005 Ford Escape HEVs gets over 25 miles per gallon on the highway?

(b) What proportion of automatic four-wheel-drive 2005 Ford Escape HEVs gets less than 30 miles per gallon on the highway?

(c) What is the probability that a randomly selected automatic four-wheel-drive 2005 Ford Escape HEV gets between 26 and 34 miles per gallon on the highway?

(d) What is the probability that a randomly selected automatic four-wheel-drive 2005 Ford Escape HEV gets over 35 miles per gallon on the highway?

(e) What is the percentile rank of an automatic four-wheel-drive 2005 Ford Escape HEV that gets 32 miles per gallon?

(f) What is the percentile rank of an automatic four-wheel-drive 2005 Ford Escape HEV that gets 25 miles per gallon?

36. Creative Thinking According to a *USA Today* "Snapshot," 20% of adults surveyed do their most creative thinking while driving. Suppose you conduct a survey of 250 adults and find that 30 do their most creative thinking while driving.

(a) Compute the probability that, in a random sample of 250 adults, 30 or fewer do their most creative thinking while driving.

(b) Does this result contradict the *USA Today* "Snapshot"? Explain.

37. Suppose a continuous random variable X is uniformly distributed with $0 \leq X \leq 20$.

(a) Draw a graph of the uniform density function.

(b) What is $P(0 \leq X \leq 5)$?

(c) What is $P(10 \leq X \leq 18)$?

38. List the properties of the standard normal curve.

39. Explain how to use a normal probability plot to assess normality.

THE CHAPTER 7 CASE STUDY IS LOCATED ON THE CD THAT ACCOMPANIES THIS TEXT.