CHAPTER 6

Continuous Probability Distributions

CONTENTS

STATISTICS *in* PRACTICE

PROCTER & GAMBLE*
CINCINNATI, OHIO

Procter & Gamble (P&G) produces and markets such products as detergents, disposable diapers, over-the-counter pharmaceuticals, dentifrices, bar soaps, mouthwashes, and paper towels. Worldwide, it has the leading brand in more categories than any other consumer products company. Since its merger with Gillette, P&G also produces and markets razors, blades, and many other personal care products.

As a leader in the application of statistical methods in decision making, P&G employs people with diverse academic backgrounds: engineering, statistics, operations research, and business. The major quantitative technologies for which these people provide support are probabilistic decision and risk analysis, advanced simulation, quality improvement, and quantitative methods (e.g., linear programming, regression analysis, probability analysis).

The Industrial Chemicals Division of P&G is a major supplier of fatty alcohols derived from natural substances such as coconut oil and from petroleum-based derivatives. The division wanted to know the economic risks and opportunities of expanding its fatty-alcohol production facilities, so it called in P&G's experts in probabilistic decision and risk analysis to help. After structuring and modeling the problem, they determined that the key to profitability was the cost difference between the petroleum- and coconut-based raw materials. Future costs were unknown, but the analysts were able to approximate them with the following continuous random variables.

x = the coconut oil price per pound of fatty alcohol

and

y = the petroleum raw material price per pound of fatty alcohol

Because the key to profitability was the difference between these two random variables, a third random variable, $d = x - y$, was used in the analysis. Experts were interviewed to determine the probability distributions for x and y. In turn, this information was used to develop a probability distribution for the difference in prices d. This continuous probability distribution showed a .90 probability that the price difference would be \$.0655 or less and a .50 probability that the price difference would be \$.035 or less. In addition, there was only a .10 probability that the price difference would be \$.0045 or less.[†]

Some of Procter & Gamble's many well-known products. © Robert Sullivan/AFP/Getty Images.

The Industrial Chemicals Division thought that being able to quantify the impact of raw material price differences was key to reaching a consensus. The probabilities obtained were used in a sensitivity analysis of the raw material price difference. The analysis yielded sufficient insight to form the basis for a recommendation to management.

The use of continuous random variables and their probability distributions was helpful to P&G in analyzing the economic risks associated with its fatty-alcohol production. In this chapter, you will gain an understanding of continuous random variables and their probability distributions, including one of the most important probability distributions in statistics, the normal distribution.

*The authors are indebted to Joel Kahn of Procter & Gamble for providing this Statistics in Practice.

[†]The price differences stated here have been modified to protect proprietary data.

In the preceding chapter we discussed discrete random variables and their probability distributions. In this chapter we turn to the study of continuous random variables. Specifically, we discuss three continuous probability distributions: the uniform, the normal, and the exponential.

A fundamental difference separates discrete and continuous random variables in terms of how probabilities are computed. For a discrete random variable, the probability function $f(x)$ provides the probability that the random variable assumes a particular value. With continuous random variables, the counterpart of the probability function is the **probability density function**, also denoted by $f(x)$. The difference is that the probability density function does not directly provide probabilities. However, the area under the graph of $f(x)$ corresponding to a given interval does provide the probability that the continuous random variable x assumes a value in that interval. So when we compute probabilities for continuous random variables we are computing the probability that the random variable assumes any value in an interval.

Because the area under the graph of $f(x)$ at any particular point is zero, one of the implications of the definition of probability for continuous random variables is that the probability of any particular value of the random variable is zero. In Section 6.1 we demonstrate these concepts for a continuous random variable that has a uniform distribution.

Much of the chapter is devoted to describing and showing applications of the normal distribution. The normal distribution is of major importance because of its wide applicability and its extensive use in statistical inference. The chapter closes with a discussion of the exponential distribution. The exponential distribution is useful in applications involving such factors as waiting times and service times.

6.1 Uniform Probability Distribution

Whenever the probability is proportional to the length of the interval, the random variable is uniformly distributed.

Consider the random variable x representing the flight time of an airplane traveling from Chicago to New York. Suppose the flight time can be any value in the interval from 120 minutes to 140 minutes. Because the random variable x can assume any value in that interval, x is a continuous rather than a discrete random variable. Let us assume that sufficient actual flight data are available to conclude that the probability of a flight time within any 1-minute interval is the same as the probability of a flight time within any other 1-minute interval contained in the larger interval from 120 to 140 minutes. With every 1-minute interval being equally likely, the random variable x is said to have a **uniform probability distribution**. The probability density function, which defines the uniform distribution for the flight-time random variable, is

$$f(x) = \begin{cases} 1/20 & \text{for } 120 \le x \le 140 \\ 0 & \text{elsewhere} \end{cases}$$

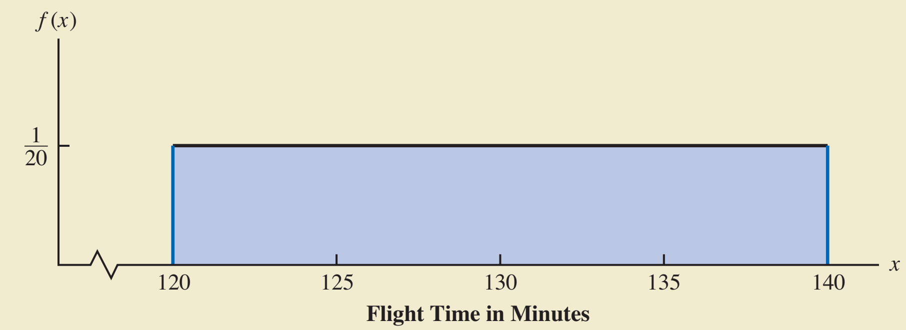
Figure 6.1 is a graph of this probability density function. In general, the uniform probability density function for a random variable x is defined by the following formula.

UNIFORM PROBABILITY DENSITY FUNCTION

$$f(x) = \begin{cases} \dfrac{1}{b - a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere} \end{cases} \tag{6.1}$$

For the flight-time random variable, $a = 120$ and $b = 140$.

FIGURE 6.1 UNIFORM PROBABILITY DISTRIBUTION FOR FLIGHT TIME

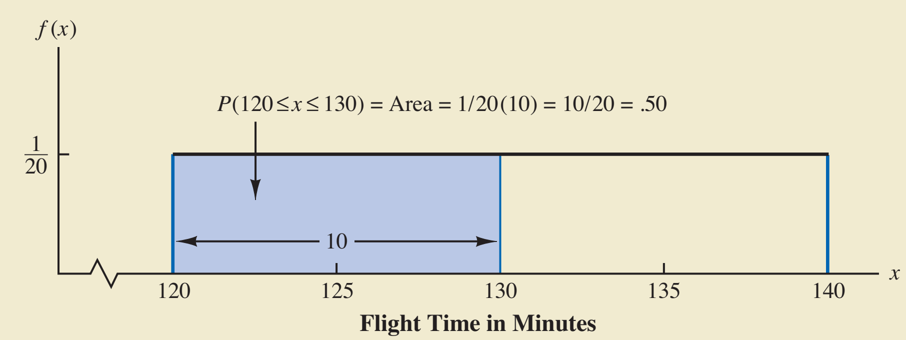


As noted in the introduction, for a continuous random variable, we consider probability only in terms of the likelihood that a random variable assumes a value within a specified interval. In the flight time example, an acceptable probability question is: What is the probability that the flight time is between 120 and 130 minutes? That is, what is $P(120 \leq x \leq 130)$? Because the flight time must be between 120 and 140 minutes and because the probability is described as being uniform over this interval, we feel comfortable saying $P(120 \leq x \leq 130) = .50$. In the following subsection we show that this probability can be computed as the area under the graph of $f(x)$ from 120 to 130 (see Figure 6.2).

Area as a Measure of Probability

Let us make an observation about the graph in Figure 6.2. Consider the area under the graph of $f(x)$ in the interval from 120 to 130. The area is rectangular, and the area of a rectangle is simply the width multiplied by the height. With the width of the interval equal to $130 - 120 = 10$ and the height equal to the value of the probability density function $f(x) = 1/20$, we have area = width $\times$ height = $10(1/20) = 10/20 = .50$.

FIGURE 6.2 AREA PROVIDES PROBABILITY OF A FLIGHT TIME BETWEEN 120 AND 130 MINUTES



What observation can you make about the area under the graph of $f(x)$ and probability? They are identical! Indeed, this observation is valid for all continuous random variables. Once a probability density function $f(x)$ is identified, the probability that x takes a value between some lower value x_1 and some higher value x_2 can be found by computing the area under the graph of $f(x)$ over the interval from x_1 to x_2.

Given the uniform distribution for flight time and using the interpretation of area as probability, we can answer any number of probability questions about flight times. For example, what is the probability of a flight time between 128 and 136 minutes? The width of the interval is $136 - 128 = 8$. With the uniform height of $f(x) = 1/20$, we see that $P(128 \leq x \leq 136) = 8(1/20) = .40$.

Note that $P(120 \leq x \leq 140) = 20(1/20) = 1$; that is, the total area under the graph of $f(x)$ is equal to 1. This property holds for all continuous probability distributions and is the analog of the condition that the sum of the probabilities must equal 1 for a discrete probability function. For a continuous probability density function, we must also require that $f(x) \geq 0$ for all values of x. This requirement is the analog of the requirement that $f(x) \geq 0$ for discrete probability functions.

Two major differences stand out between the treatment of continuous random variables and the treatment of their discrete counterparts.

1. We no longer talk about the probability of the random variable assuming a particular value. Instead, we talk about the probability of the random variable assuming a value within some given interval.
2. The probability of a continuous random variable assuming a value within some given interval from x_1 to x_2 is defined to be the area under the graph of the probability density function between x_1 and x_2. Because a single point is an interval of zero width, this implies that the probability of a continuous random variable assuming any particular value exactly is zero. It also means that the probability of a continuous random variable assuming a value in any interval is the same whether or not the endpoints are included.

To see that the probability of any single point is 0, refer to Figure 6.2 and compute the probability of a single point, say, $x = 125$. $P(x = 125) = P(125 \leq x \leq 125) = 0(1/20) = 0$.

The calculation of the expected value and variance for a continuous random variable is analogous to that for a discrete random variable. However, because the computational procedure involves integral calculus, we leave the derivation of the appropriate formulas to more advanced texts.

For the uniform continuous probability distribution introduced in this section, the formulas for the expected value and variance are

$$E(x) = \frac{a + b}{2}$$

$$\text{Var}(x) = \frac{(b - a)^2}{12}$$

In these formulas, a is the smallest value and b is the largest value that the random variable may assume.

Applying these formulas to the uniform distribution for flight times from Chicago to New York, we obtain

$$E(x) = \frac{(120 + 140)}{2} = 130$$

$$\text{Var}(x) = \frac{(140 - 120)^2}{12} = 33.33$$

The standard deviation of flight times can be found by taking the square root of the variance. Thus, $\sigma = 5.77$ minutes.

NOTES AND COMMENTS

To see more clearly why the height of a probability density function is not a probability, think about a random variable with the following uniform probability distribution.

$$f(x) = \begin{cases} 2 & \text{for } 0 \le x \le .5 \\ 0 & \text{elsewhere} \end{cases}$$

The height of the probability density function, $f(x)$, is 2 for values of x between 0 and .5. However, we know probabilities can never be greater than 1. Thus, we see that $f(x)$ cannot be interpreted as the probability of x.

Exercises

Methods

SELF test

1. The random variable x is known to be uniformly distributed between 1.0 and 1.5.
   a. Show the graph of the probability density function.
   b. Compute $P(x = 1.25)$.
   c. Compute $P(1.0 \le x \le 1.25)$.
   d. Compute $P(1.20 < x < 1.5)$.
2. The random variable x is known to be uniformly distributed between 10 and 20.
   a. Show the graph of the probability density function.
   b. Compute $P(x < 15)$.
   c. Compute $P(12 \le x \le 18)$.
   d. Compute $E(x)$.
   e. Compute Var(x).

Applications

3. Delta Airlines quotes a flight time of 2 hours, 5 minutes for its flights from Cincinnati to Tampa. Suppose we believe that actual flight times are uniformly distributed between 2 hours and 2 hours, 20 minutes.
   a. Show the graph of the probability density function for flight time.
   b. What is the probability that the flight will be no more than 5 minutes late?
   c. What is the probability that the flight will be more than 10 minutes late?
   d. What is the expected flight time?

SELF test

4. Most computer languages include a function that can be used to generate random numbers. In Excel, the RAND function can be used to generate random numbers between 0 and 1. If we let x denote a random number generated using RAND, then x is a continuous random variable with the following probability density function.

$$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

   a. Graph the probability density function.
   b. What is the probability of generating a random number between .25 and .75?
   c. What is the probability of generating a random number with a value less than or equal to .30?
   d. What is the probability of generating a random number with a value greater than .60?
   e. Generate 50 random numbers by entering =RAND() into 50 cells of an Excel worksheet.
   f. Compute the mean and standard deviation for the random numbers in part (e).

5. The driving distance for the top 100 golfers on the PGA tour is between 284.7 and 310.6 yards (*Golfweek,* March 29, 2003). Assume that the driving distance for these golfers is uniformly distributed over this interval.
   a. Give a mathematical expression for the probability density function of driving distance.
   b. What is the probability the driving distance for one of these golfers is less than 290 yards?
   c. What is the probability the driving distance for one of these golfers is at least 300 yards?
   d. What is the probability the driving distance for one of these golfers is between 290 and 305 yards?
   e. How many of these golfers drive the ball at least 290 yards?
6. On average, 30-minute television sitcoms have 22 minutes of programming (CNBC, February 23, 2006). Assume that the probability distribution for minutes of programming can be approximated by a uniform distribution from 18 minutes to 26 minutes.
   a. What is the probability a sitcom will have 25 or more minutes of programming?
   b. What is the probability a sitcom will have between 21 and 25 minutes of programming?
   c. What is the probability a sitcom will have more than 10 minutes of commercials or other nonprogramming interruptions?
7. Suppose we are interested in bidding on a piece of land and we know one other bidder is interested.[1] The seller announced that the highest bid in excess of $10,000 will be accepted. Assume that the competitor's bid x is a random variable that is uniformly distributed between $10,000 and $15,000.
   a. Suppose you bid $12,000. What is the probability that your bid will be accepted?
   b. Suppose you bid $14,000. What is the probability that your bid will be accepted?
   c. What amount should you bid to maximize the probability that you get the property?
   d. Suppose you know someone who is willing to pay you $16,000 for the property. Would you consider bidding less than the amount in part (c)? Why or why not?

6.2 Normal Probability Distribution

Abraham de Moivre, a French mathematician, published The Doctrine of Chances *in 1733. He derived the normal distribution.*

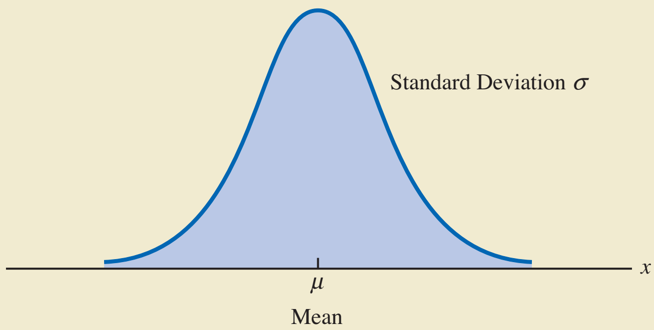
The most important probability distribution for describing a continuous random variable is the **normal probability distribution**. The normal distribution has been used in a wide variety of practical applications in which the random variables are heights and weights of people, test scores, scientific measurements, amounts of rainfall, and other similar values. It is also widely used in statistical inference, which is the major topic of the remainder of this book. In such applications, the normal distribution provides a description of the likely results obtained through sampling.

Normal Curve

The form, or shape, of the normal distribution is illustrated by the bell-shaped normal curve in Figure 6.3. The probability density function that defines the bell-shaped curve of the normal distribution follows.

[1] This exercise is based on a problem suggested to us by Professor Roger Myerson of Northwestern University.

FIGURE 6.3 BELL-SHAPED CURVE FOR THE NORMAL DISTRIBUTION



NORMAL PROBABILITY DENSITY FUNCTION

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad (6.2)$$

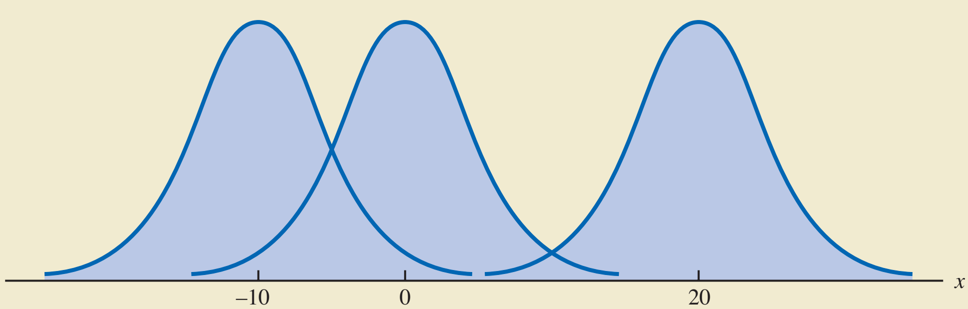
where

μ = mean
σ = standard deviation
π = 3.14159
e = 2.71828

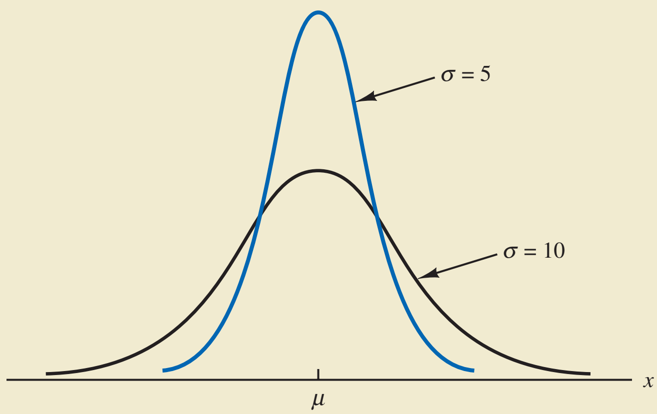
The normal curve has two parameters, μ and σ. They determine the location and shape of the normal distribution.

We make several observations about the characteristics of the normal distribution.

1. The entire family of normal distributions is differentiated by two parameters: the mean μ and the standard deviation σ.
2. The highest point on the normal curve is at the mean, which is also the median and mode of the distribution.
3. The mean of the distribution can be any numerical value: negative, zero, or positive. Three normal distributions with the same standard deviation but three different means (−10, 0, and 20) are shown here.



4. The normal distribution is symmetric, with the shape of the normal curve to the left of the mean a mirror image of the shape of the normal curve to the right of the mean. The tails of the normal curve extend to infinity in both directions and theoretically never touch the horizontal axis. Because it is symmetric, the normal distribution is not skewed; its skewness measure is zero.
5. The standard deviation determines how flat and wide the normal curve is. Larger values of the standard deviation result in wider, flatter curves, showing more variability in the data. Two normal distributions with the same mean but with different standard deviations are shown here.



6. Probabilities for the normal random variable are given by areas under the normal curve. The total area under the curve for the normal distribution is 1. Because the distribution is symmetric, the area under the curve to the left of the mean is .50 and the area under the curve to the right of the mean is .50.
7. The percentage of values in some commonly used intervals are
   a. 68.3% of the values of a normal random variable are within plus or minus one standard deviation of its mean.
   b. 95.4% of the values of a normal random variable are within plus or minus two standard deviations of its mean.
   c. 99.7% of the values of a normal random variable are within plus or minus three standard deviations of its mean.

These percentages are the basis for the empirical rule introduced in Section 3.3.

Figure 6.4 shows properties (a), (b), and (c) graphically.

Standard Normal Probability Distribution

A random variable that has a normal distribution with a mean of zero and a standard deviation of one is said to have a **standard normal probability distribution**. The letter z is commonly used to designate this particular normal random variable. Figure 6.5 is the graph of the standard normal distribution. It has the same general appearance as other normal distributions, but with the special properties of $\mu = 0$ and $\sigma = 1$.

FIGURE 6.4 AREAS UNDER THE CURVE FOR ANY NORMAL DISTRIBUTION

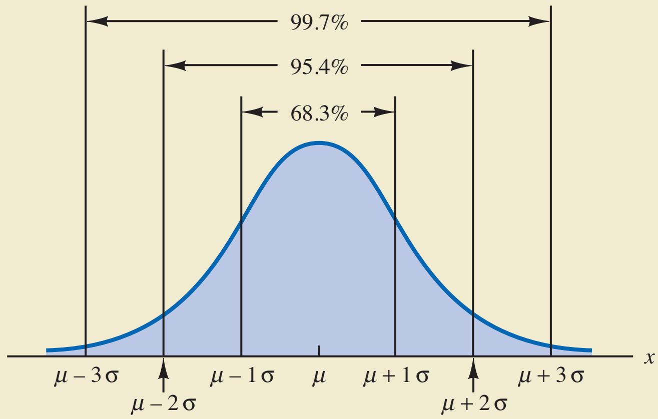
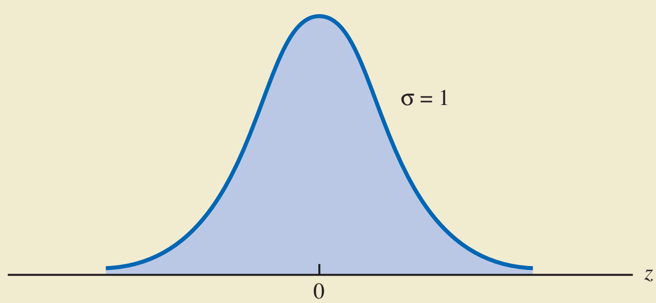


FIGURE 6.5 THE STANDARD NORMAL DISTRIBUTION



Because $\mu = 0$ and $\sigma = 1$, the formula for the standard normal probability density function is a simpler version of equation (6.2).

STANDARD NORMAL DENSITY FUNCTION

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

As with other continuous random variables, probability calculations with any normal distribution are made by computing areas under the graph of the probability density function. Thus, to find the probability that a normal random variable is within any specific interval, we must compute the area under the normal curve over that interval.

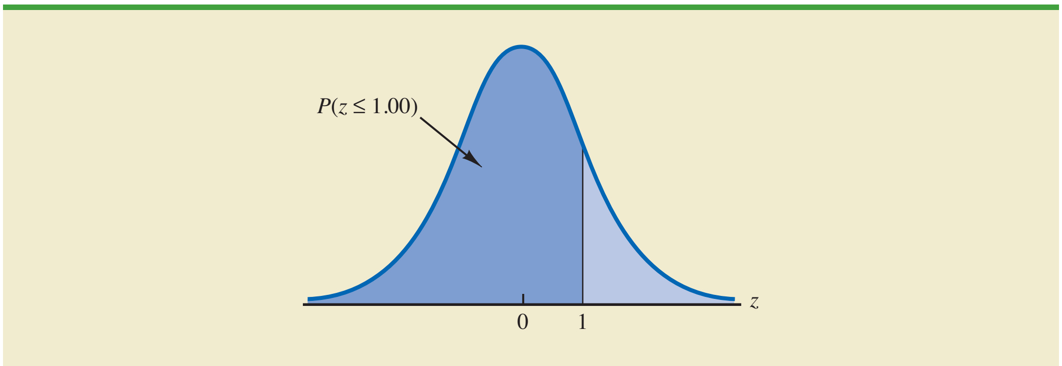
For the normal probability density function, the height of the normal curve varies and more advanced mathematics is required to compute the areas that represent probability.

For the standard normal distribution, areas under the normal curve have been computed and are available in tables that can be used to compute probabilities. Such a table appears on the two pages inside the front cover of the text. The table on the left-hand page contains areas, or cumulative probabilities, for z values less than or equal to the mean of zero. The table on the right-hand page contains areas, or cumulative probabilities, for z values greater than or equal to the mean of zero.

The three types of probabilities we need to compute include (1) the probability that the standard normal random variable z will be less than or equal to a given value; (2) the probability that z will be between two given values; and (3) the probability that z will be greater than or equal to a given value. To see how the cumulative probability table for the standard normal distribution can be used to compute these three types of probabilities, let us consider some examples.

Because the standard normal random variable is continuous, $P(z \leq 1.00) = P(z < 1.00)$.

We start by showing how to compute the probability that z is less than or equal to 1.00; that is, $P(z \leq 1.00)$. This cumulative probability is the area under the normal curve to the left of $z = 1.00$ in the following graph.

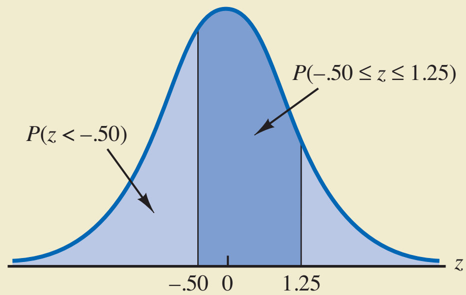


Refer to the right-hand page of the standard normal probability table inside the front cover of the text. The cumulative probability corresponding to $z = 1.00$ is the table value located at the intersection of the row labeled 1.0 and the column labeled .00. First we find 1.0 in the left column of the table and then find .00 in the top row of the table. By looking in the body of the table, we find that the 1.0 row and the .00 column intersect at the value of .8413; thus, $P(z \leq 1.00) = .8413$. The following excerpt from the probability table shows these steps.

z	**.00**	**.01**	**.02**
.			
.			
.			
.9	.8159	.8186	.8212
1.0	**.8413**	.8438	.8461
1.1	.8643	.8665	.8686
1.2	.8849	.8869	.8888
.			
.			
.			

$P(z \leq 1.00)$

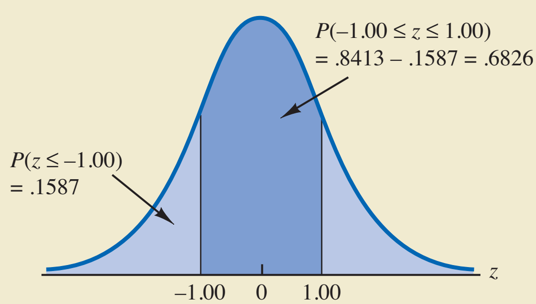
To illustrate the second type of probability calculation we show how to compute the probability that z is in the interval between $-.50$ and 1.25; that is, $P(-.50 \leq z \leq 1.25)$. The following graph shows this area, or probability.



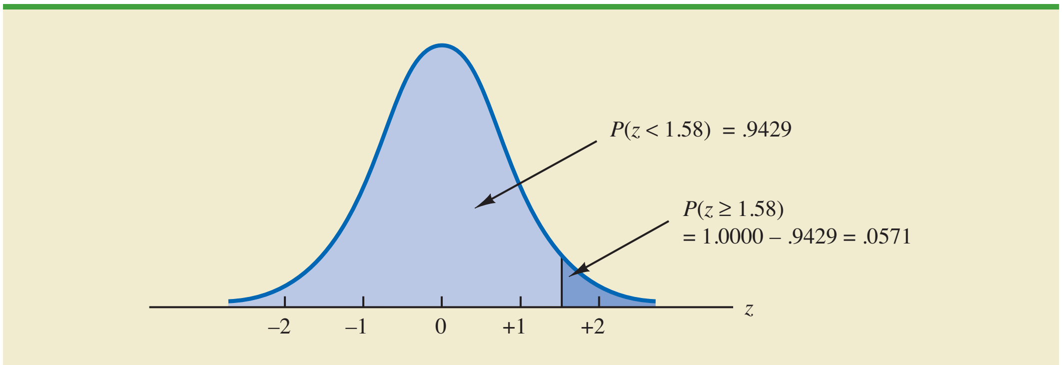
Three steps are required to compute this probability. First, we find the area under the normal curve to the left of $z = 1.25$. Second, we find the area under the normal curve to the left of $z = -.50$. Finally, we subtract the area to the left of $z = -.50$ from the area to the left of $z = 1.25$ to find $P(-.50 \le z \le 1.25)$.

To find the area under the normal curve to the left of $z = 1.25$, we first locate the 1.2 row in the standard normal probability table and then move across to the .05 column. Because the table value in the 1.2 row and the .05 column is .8944, $P(z \le 1.25) = .8944$. Similarly, to find the area under the curve to the left of $z = -.50$, we use the left-hand page of the table to locate the table value in the $-.5$ row and the .00 column; with a table value of .3085, $P(z \le -.50) = .3085$. Thus, $P(-.50 \le z \le 1.25) = P(z \le 1.25) - P(z \le -.50) = .8944 - .3085 = .5859$.

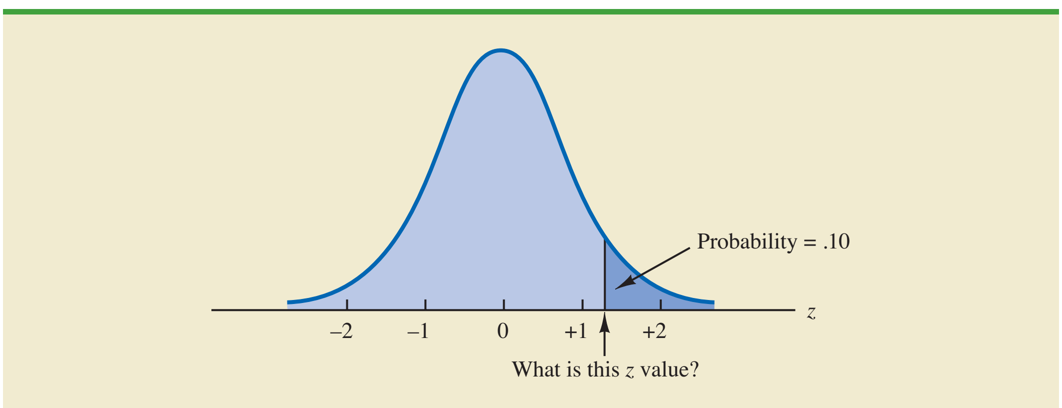
Let us consider another example of computing the probability that z is in the interval between two given values. Often it is of interest to compute the probability that a normal random variable assumes a value within a certain number of standard deviations of the mean. Suppose we want to compute the probability that the standard normal random variable is within one standard deviation of the mean; that is, $P(-1.00 \le z \le 1.00)$. To compute this probability we must find the area under the curve between -1.00 and 1.00. Earlier we found that $P(z \le 1.00) = .8413$. Referring again to the table inside the front cover of the book, we find that the area under the curve to the left of $z = -1.00$ is .1587, so $P(z \le -1.00) = .1587$. Therefore, $P(-1.00 \le z \le 1.00) = P(z \le 1.00) - P(z \le -1.00) = .8413 - .1587 = .6826$. This probability is shown graphically in the following figure.



To illustrate how to make the third type of probability computation, suppose we want to compute the probability of obtaining a z value of at least 1.58; that is, $P(z \geq 1.58)$. The value in the $z = 1.5$ row and the .08 column of the cumulative normal table is .9429; thus, $P(z < 1.58) = .9429$. However, because the total area under the normal curve is 1, $P(z \geq 1.58) = 1 - .9429 = .0571$. This probability is shown in the following figure.



In the preceding illustrations, we showed how to compute probabilities given specified z values. In some situations, we are given a probability and are interested in working backward to find the corresponding z value. Suppose we want to find a z value such that the probability of obtaining a larger z value is .10. The following figure shows this situation graphically.



Given a probability, we can use the standard normal table in an inverse fashion to find the corresponding z value.

This problem is the inverse of those in the preceding examples. Previously, we specified the z value of interest and then found the corresponding probability, or area. In this example, we are given the probability, or area, and asked to find the corresponding z value. To do so, we use the standard normal probability table somewhat differently.

Recall that the standard normal probability table gives the area under the curve to the left of a particular z value. We have been given the information that the area in the upper tail of the curve is .10. Hence, the area under the curve to the left of the unknown z value must equal .9000. Scanning the body of the table, we find .8997 is the cumulative probability value closest to .9000. The section of the table providing this result follows.

z	**.06**	**.07**	**.08**	**.09**
.				
.				
.				
1.0	.8554	.8577	.8599	.8621
1.1	.8770	.8790	.8810	.8830
1.2	.8962	.8980	.8997	.9015
1.3	.9131	.9147	.9162	.9177
1.4	.9279	.9292	.9306	.9319
.				
.				
.				

Cumulative probability value closest to .9000

Reading the z value from the left-most column and the top row of the table, we find that the corresponding z value is 1.28. Thus, an area of approximately .9000 (actually .8997) will be to the left of $z = 1.28$.[2] In terms of the question originally asked, there is an approximately .10 probability of a z value larger than 1.28.

The examples illustrate that the table of cumulative probabilities for the standard normal probability distribution can be used to find probabilities associated with values of the standard normal random variable z. Two types of questions can be asked. The first type of question specifies a value, or values, for z and asks us to use the table to determine the corresponding areas or probabilities. The second type of question provides an area, or probability, and asks us to use the table to determine the corresponding z value. Thus, we need to be flexible in using the standard normal probability table to answer the desired probability question. In most cases, sketching a graph of the standard normal probability distribution and shading the appropriate area will help to visualize the situation and aid in determining the correct answer.

Computing Probabilities for Any Normal Probability Distribution

The reason for discussing the standard normal distribution so extensively is that probabilities for all normal distributions are computed by using the standard normal distribution. That is, when we have a normal distribution with any mean μ and any standard deviation σ, we answer probability questions about the distribution by first converting to the standard normal distribution. Then we can use the standard normal probability table and the appropriate z values to find the desired probabilities. The formula used to convert any normal random variable x with mean μ and standard deviation σ to the standard normal random variable z follows.

The formula for the standard normal random variable is similar to the formula we introduced in Chapter 3 for computing z-scores for a data set.

CONVERTING TO THE STANDARD NORMAL RANDOM VARIABLE

$$z = \frac{x - \mu}{\sigma} \tag{6.3}$$

[2] We could use interpolation in the body of the table to get a better approximation of the z value that corresponds to an area of .9000. Doing so to provide one more decimal place of accuracy would yield a z value of 1.282. However, in most practical situations, sufficient accuracy is obtained by simply using the table value closest to the desired probability.

A value of x equal to its mean μ results in $z = (\mu - \mu)/\sigma = 0$. Thus, we see that a value of x equal to its mean μ corresponds to $z = 0$. Now suppose that x is one standard deviation above its mean; that is, $x = \mu + \sigma$. Applying equation (6.3), we see that the corresponding z value is $z = [(\mu + \sigma) - \mu]/\sigma = \sigma/\sigma = 1$. Thus, an x value that is one standard deviation above its mean corresponds to $z = 1$. In other words, *we can interpret z as the number of standard deviations that the normal random variable x is from its mean μ.*

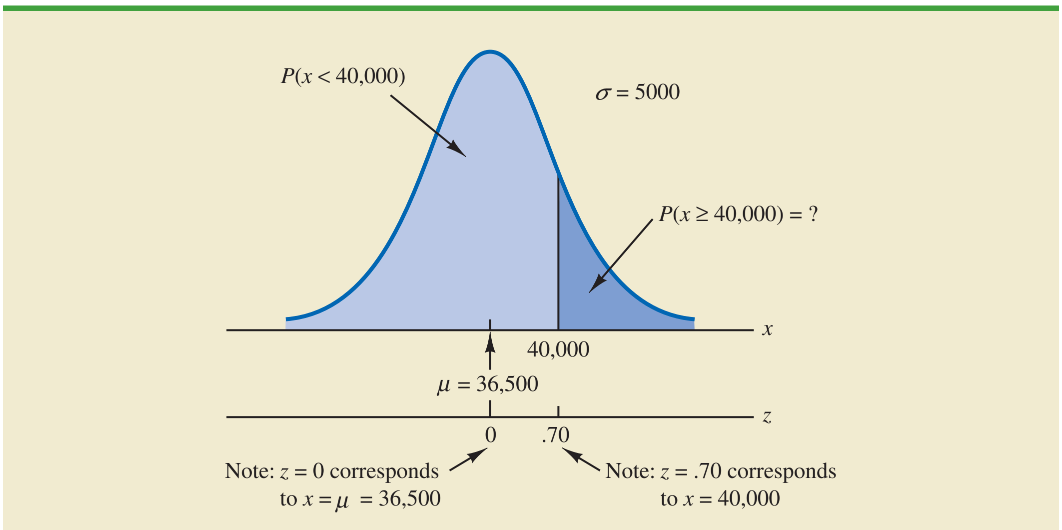
To see how this conversion enables us to compute probabilities for any normal distribution, suppose we have a normal distribution with $\mu = 10$ and $\sigma = 2$. What is the probability that the random variable x is between 10 and 14? Using equation (6.3), we see that at $x = 10$, $z = (x - \mu)/\sigma = (10 - 10)/2 = 0$ and that at $x = 14$, $z = (14 - 10)/2 = 4/2 = 2$. Thus, the answer to our question about the probability of x being between 10 and 14 is given by the equivalent probability that z is between 0 and 2 for the standard normal distribution. In other words, the probability that we are seeking is the probability that the random variable x is between its mean and two standard deviations above the mean. Using $z = 2.00$ and the standard normal probability table inside the front cover of the text, we see that $P(z \leq 2) = .9772$. Because $P(z \leq 0) = .5000$, we can compute $P(.00 \leq z \leq 2.00) = P(z \leq 2) - P(z \leq 0) = .9772 - .5000 = .4772$. Hence the probability that x is between 10 and 14 is .4772.

Grear Tire Company Problem

We turn now to an application of the normal probability distribution. Suppose the Grear Tire Company developed a new steel-belted radial tire to be sold through a national chain of discount stores. Because the tire is a new product, Grear's managers believe that the mileage guarantee offered with the tire will be an important factor in the acceptance of the product. Before finalizing the tire mileage guarantee policy, Grear's managers want probability information about x = number of miles the tires will last.

From actual road tests with the tires, Grear's engineering group estimated that the mean tire mileage is $\mu = 36{,}500$ miles and that the standard deviation is $\sigma = 5000$. In addition, the data collected indicate that a normal distribution is a reasonable assumption. What percentage of the tires can be expected to last more than 40,000 miles? In other words, what is the probability that the tire mileage, x, will exceed 40,000? This question can be answered by finding the area of the darkly shaded region in Figure 6.6.

FIGURE 6.6 GREAR TIRE COMPANY MILEAGE DISTRIBUTION



At $x = 40{,}000$, we have

$$z = \frac{x - \mu}{\sigma} = \frac{40{,}000 - 36{,}500}{5000} = \frac{3500}{5000} = .70$$

Refer now to the bottom of Figure 6.6. We see that a value of $x = 40{,}000$ on the Grear Tire normal distribution corresponds to a value of $z = .70$ on the standard normal distribution. Using the standard normal probability table, we see that the area under the standard normal curve to the left of $z = .70$ is .7580. Thus, $1.000 - .7580 = .2420$ is the probability that z will exceed .70 and hence x will exceed 40,000. We can conclude that about 24.2% of the tires will exceed 40,000 in mileage.

Let us now assume that Grear is considering a guarantee that will provide a discount on replacement tires if the original tires do not provide the guaranteed mileage. What should the guarantee mileage be if Grear wants no more than 10% of the tires to be eligible for the discount guarantee? This question is interpreted graphically in Figure 6.7.

According to Figure 6.7, the area under the curve to the left of the unknown guarantee mileage must be .10. So, we must first find the z-value that cuts off an area of .10 in the left tail of a standard normal distribution. Using the standard normal probability table, we see that $z = -1.28$ cuts off an area of .10 in the lower tail. Hence, $z = -1.28$ is the value of the standard normal random variable corresponding to the desired mileage guarantee on the Grear Tire normal distribution. To find the value of x corresponding to $z = -1.28$, we have

The guarantee mileage we need to find is 1.28 standard deviations below the mean. Thus, $x = \mu - 1.28\sigma$.

$$z = \frac{x - \mu}{\sigma} = -1.28$$
$$x - \mu = -1.28\sigma$$
$$x = \mu - 1.28\sigma$$

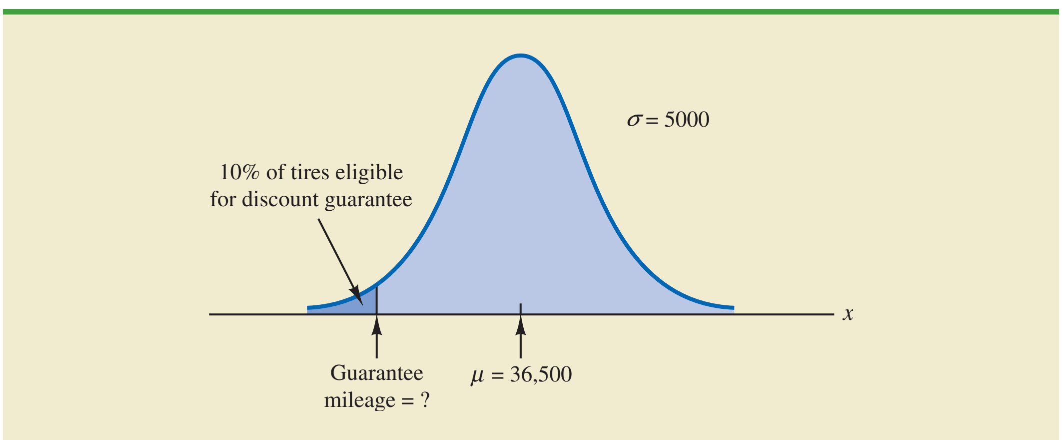
With $\mu = 36{,}500$ and $\sigma = 5000$,

$$x = 36{,}500 - 1.28(5000) = 30{,}100$$

With the guarantee set at 30,000 miles, the actual percentage eligible for the guarantee will be 9.68%.

Thus, a guarantee of 30,100 miles will meet the requirement that approximately 10% of the tires will be eligible for the guarantee. Perhaps, with this information, the firm will set its tire mileage guarantee at 30,000 miles.

FIGURE 6.7 GREAR'S DISCOUNT GUARANTEE



Again, we see the important role that probability distributions play in providing decision-making information. Namely, once a probability distribution is established for a particular application, it can be used to obtain probability information about the problem. Probability does not make a decision recommendation directly, but it provides information that helps the decision maker better understand the risks and uncertainties associated with the problem. Ultimately, this information may assist the decision maker in reaching a good decision.

EXERCISES

Methods

8. Using Figure 6.4 as a guide, sketch a normal curve for a random variable x that has a mean of $\mu = 100$ and a standard deviation of $\sigma = 10$. Label the horizontal axis with values of 70, 80, 90, 100, 110, 120, and 130.
9. A random variable is normally distributed with a mean of $\mu = 50$ and a standard deviation of $\sigma = 5$.
   a. Sketch a normal curve for the probability density function. Label the horizontal axis with values of 35, 40, 45, 50, 55, 60, and 65. Figure 6.4 shows that the normal curve almost touches the horizontal axis at three standard deviations below and at three standard deviations above the mean (in this case at 35 and 65).
   b. What is the probability the random variable will assume a value between 45 and 55?
   c. What is the probability the random variable will assume a value between 40 and 60?
10. Draw a graph for the standard normal distribution. Label the horizontal axis at values of -3, -2, -1, 0, 1, 2, and 3. Then use the table of probabilities for the standard normal distribution inside the front cover of the text to compute the following probabilities.
    a. $P(z \leq 1.5)$
    b. $P(z \leq 1)$
    c. $P(1 \leq z \leq 1.5)$
    d. $P(0 < z < 2.5)$
11. Given that z is a standard normal random variable, compute the following probabilities.
    a. $P(z \leq -1.0)$
    b. $P(z \geq -1)$
    c. $P(z \geq -1.5)$
    d. $P(-2.5 \leq z)$
    e. $P(-3 < z \leq 0)$
12. Given that z is a standard normal random variable, compute the following probabilities.
    a. $P(0 \leq z \leq .83)$
    b. $P(-1.57 \leq z \leq 0)$
    c. $P(z > .44)$
    d. $P(z \geq -.23)$
    e. $P(z < 1.20)$
    f. $P(z \leq -.71)$

SELF test

13. Given that z is a standard normal random variable, compute the following probabilities.
    a. $P(-1.98 \leq z \leq .49)$
    b. $P(.52 \leq z \leq 1.22)$
    c. $P(-1.75 \leq z \leq -1.04)$
14. Given that z is a standard normal random variable, find z for each situation.
    a. The area to the left of z is .9750.
    b. The area between 0 and z is .4750.
    c. The area to the left of z is .7291.
    d. The area to the right of z is .1314.
    e. The area to the left of z is .6700.
    f. The area to the right of z is .3300.

SELF test

15. Given that z is a standard normal random variable, find z for each situation.
    a. The area to the left of z is .2119.
    b. The area between $-z$ and z is .9030.
    c. The area between $-z$ and z is .2052.
    d. The area to the left of z is .9948.
    e. The area to the right of z is .6915.
16. Given that z is a standard normal random variable, find z for each situation.
    a. The area to the right of z is .01.
    b. The area to the right of z is .025.
    c. The area to the right of z is .05.
    d. The area to the right of z is .10.

Applications

17. For borrowers with good credit scores, the mean debt for revolving and installment accounts is \$15,015 (*BusinessWeek,* March 20, 2006). Assume the standard deviation is \$3540 and that debt amounts are normally distributed.
    a. What is the probability that the debt for a borrower with good credit is more than \$18,000?
    b. What is the probability that the debt for a borrower with good credit is less than \$10,000?
    c. What is the probability that the debt for a borrower with good credit is between \$12,000 and \$18,000?
    d. What is the probability that the debt for a borrower with good credit is no more than \$14,000?

SELF test

18. The average stock price for companies making up the S&P 500 is \$30, and the standard deviation is \$8.20 (*BusinessWeek,* Special Annual Issue, Spring 2003). Assume the stock prices are normally distributed.
    a. What is the probability a company will have a stock price of at least \$40?
    b. What is the probability a company will have a stock price no higher than \$20?
    c. How high does a stock price have to be to put a company in the top 10%?
19. In an article about the cost of health care, *Money* magazine reported that a visit to a hospital emergency room for something as simple as a sore throat has a mean cost of \$328 (*Money,* January 2009). Assume that the cost for this type of hospital emergency room visit is normally distributed with a standard deviation of \$92. Answer the following questions about the cost of a hospital emergency room visit for this medical service.
    a. What is the probability that the cost will be more than \$500?
    b. What is the probability that the cost will be less than \$250?
    c. What is the probability that the cost will be between \$300 and \$400?
    d. If the cost to a patient is in the lower 8% of charges for this medical service, what was the cost of this patient's emergency room visit?
20. In January 2003, the American worker spent an average of 77 hours logged on to the Internet while at work (CNBC, March 15, 2003). Assume the population mean is 77 hours, the times are normally distributed, and that the standard deviation is 20 hours.
    a. What is the probability that in January 2003 a randomly selected worker spent fewer than 50 hours logged on to the Internet?
    b. What percentage of workers spent more than 100 hours in January 2003 logged on to the Internet?
    c. A person is classified as a heavy user if he or she is in the upper 20% of usage. In January 2003, how many hours did a worker have to be logged on to the Internet to be considered a heavy user?
21. A person must score in the upper 2% of the population on an IQ test to qualify for membership in Mensa, the international high-IQ society (*U.S. Airways Attaché,* September 2000). If IQ scores are normally distributed with a mean of 100 and a standard deviation of 15, what score must a person have to qualify for Mensa?

22. The mean hourly pay rate for financial managers in the East North Central region is \$32.62, and the standard deviation is \$2.32 (Bureau of Labor Statistics, September 2005). Assume that pay rates are normally distributed.
    a. What is the probability a financial manager earns between \$30 and \$35 per hour?
    b. How high must the hourly rate be to put a financial manager in the top 10% with respect to pay?
    c. For a randomly selected financial manager, what is the probability the manager earned less than \$28 per hour?
23. The time needed to complete a final examination in a particular college course is normally distributed with a mean of 80 minutes and a standard deviation of 10 minutes. Answer the following questions.
    a. What is the probability of completing the exam in one hour or less?
    b. What is the probability that a student will complete the exam in more than 60 minutes but less than 75 minutes?
    c. Assume that the class has 60 students and that the examination period is 90 minutes in length. How many students do you expect will be unable to complete the exam in the allotted time?

WEB file
Volume

24. Trading volume on the New York Stock Exchange is heaviest during the first half hour (early morning) and last half hour (late afternoon) of the trading day. The early morning trading volumes (millions of shares) for 13 days in January and February are shown here (*Barron's,* January 23, 2006; February 13, 2006; and February 27, 2006).

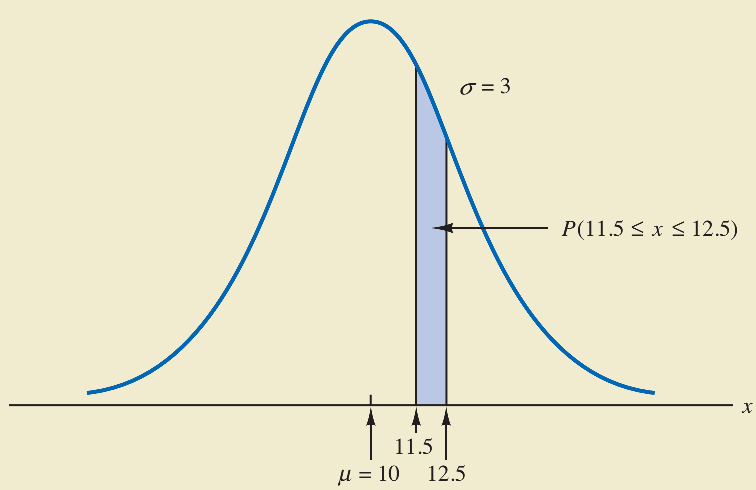
214	163	265	194	180
202	198	212	201	
174	171	211	211	

    The probability distribution of trading volume is approximately normal.
    a. Compute the mean and standard deviation to use as estimates of the population mean and standard deviation.
    b. What is the probability that, on a randomly selected day, the early morning trading volume will be less than 180 million shares?
    c. What is the probability that, on a randomly selected day, the early morning trading volume will exceed 230 million shares?
    d. How many shares would have to be traded for the early morning trading volume on a particular day to be among the busiest 5% of days?
25. According to the Sleep Foundation, the average night's sleep is 6.8 hours (*Fortune,* March 20, 2006). Assume the standard deviation is .6 hours and that the probability distribution is normal.
    a. What is the probability that a randomly selected person sleeps more than 8 hours?
    b. What is the probability that a randomly selected person sleeps 6 hours or less?
    c. Doctors suggest getting between 7 and 9 hours of sleep each night. What percentage of the population gets this much sleep?

6.3 Normal Approximation of Binomial Probabilities

In Section 5.4 we presented the discrete binomial distribution. Recall that a binomial experiment consists of a sequence of n identical independent trials with each trial having two possible outcomes, a success or a failure. The probability of a success on a trial is the same for all trials and is denoted by p. The binomial random variable is the number of successes in the n trials, and probability questions pertain to the probability of x successes in the n trials.

FIGURE 6.8 NORMAL APPROXIMATION TO A BINOMIAL PROBABILITY DISTRIBUTION WITH $n = 100$ AND $p = .10$ SHOWING THE PROBABILITY OF 12 ERRORS



When the number of trials becomes large, evaluating the binomial probability function by hand or with a calculator is difficult. In cases where $np \geq 5$, and $n(1 - p) \geq 5$, the normal distribution provides an easy-to-use approximation of binomial probabilities. When using the normal approximation to the binomial, we set $\mu = np$ and $\sigma = \sqrt{np(1 - p)}$ in the definition of the normal curve.

Let us illustrate the normal approximation to the binomial by supposing that a particular company has a history of making errors in 10% of its invoices. A sample of 100 invoices has been taken, and we want to compute the probability that 12 invoices contain errors. That is, we want to find the binomial probability of 12 successes in 100 trials. In applying the normal approximation in this case, we set $\mu = np = (100)(.1) = 10$ and $\sigma = \sqrt{np(1 - p)} = \sqrt{(100)(.1)(.9)} = 3$. A normal distribution with $\mu = 10$ and $\sigma = 3$ is shown in Figure 6.8.

Recall that, with a continuous probability distribution, probabilities are computed as areas under the probability density function. As a result, the probability of any single value for the random variable is zero. Thus to approximate the binomial probability of 12 successes, we compute the area under the corresponding normal curve between 11.5 and 12.5. The .5 that we add and subtract from 12 is called a **continuity correction factor**. It is introduced because a continuous distribution is being used to approximate a discrete distribution. Thus, $P(x = 12)$ for the *discrete* binomial distribution is approximated by $P(11.5 \leq x \leq 12.5)$ for the *continuous* normal distribution.

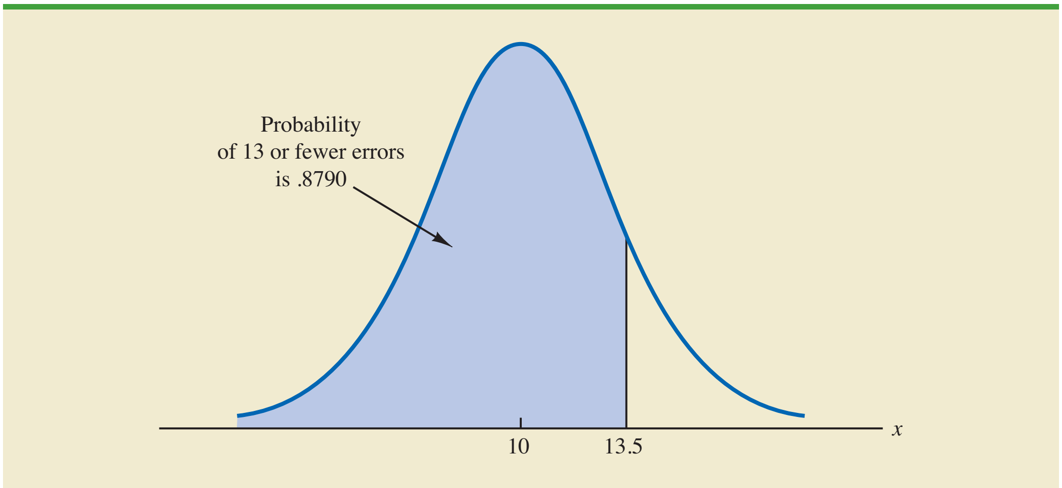
Converting to the standard normal distribution to compute $P(11.5 \leq x \leq 12.5)$, we have

$$z = \frac{x - \mu}{\sigma} = \frac{12.5 - 10.0}{3} = .83 \quad \text{at } x = 12.5$$

and

$$z = \frac{x - \mu}{\sigma} = \frac{11.5 - 10.0}{3} = .50 \quad \text{at } x = 11.5$$

FIGURE 6.9 NORMAL APPROXIMATION TO A BINOMIAL PROBABILITY DISTRIBUTION WITH $n = 100$ AND $p = .10$ SHOWING THE PROBABILITY OF 13 OR FEWER ERRORS



Using the standard normal probability table, we find that the area under the curve (in Figure 6.8) to the left of 12.5 is .7967. Similarly, the area under the curve to the left of 11.5 is .6915. Therefore, the area between 11.5 and 12.5 is .7967 − .6915 = .1052. The normal approximation to the probability of 12 successes in 100 trials is .1052.

For another illustration, suppose we want to compute the probability of 13 or fewer errors in the sample of 100 invoices. Figure 6.9 shows the area under the normal curve that approximates this probability. Note that the use of the continuity correction factor results in the value of 13.5 being used to compute the desired probability. The z value corresponding to $x = 13.5$ is

$$z = \frac{13.5 - 10.0}{3.0} = 1.17$$

The standard normal probability table shows that the area under the standard normal curve to the left of $z = 1.17$ is .8790. The area under the normal curve approximating the probability of 13 or fewer errors is given by the shaded portion of the graph in Figure 6.9.

Exercises

Methods

SELF test

26. A binomial probability distribution has $p = .20$ and $n = 100$.
    a. What are the mean and standard deviation?
    b. Is this situation one in which binomial probabilities can be approximated by the normal probability distribution? Explain.
    c. What is the probability of exactly 24 successes?
    d. What is the probability of 18 to 22 successes?
    e. What is the probability of 15 or fewer successes?
27. Assume a binomial probability distribution has $p = .60$ and $n = 200$.
    a. What are the mean and standard deviation?
    b. Is this situation one in which binomial probabilities can be approximated by the normal probability distribution? Explain.

c. What is the probability of 100 to 110 successes?
d. What is the probability of 130 or more successes?
e. What is the advantage of using the normal probability distribution to approximate the binomial probabilities? Use part (d) to explain the advantage.

Applications

SELF test

28. Although studies continue to show smoking leads to significant health problems, 20% of adults in the United States smoke. Consider a group of 250 adults.
    a. What is the expected number of adults who smoke?
    b. What is the probability that fewer than 40 smoke?
    c. What is the probability that from 55 to 60 smoke?
    d. What is the probability that 70 or more smoke?
29. An Internal Revenue Oversight Board survey found that 82% of taxpayers said that it was very important for the Internal Revenue Service (IRS) to ensure that high-income tax payers do not cheat on their tax returns (*The Wall Street Journal,* February 11, 2009).
    a. For a sample of eight taxpayers, what is the probability that at least six taxpayers say that it is very important to ensure that high-income tax payers do not cheat on their tax returns? Use the binomial distribution probability function shown in Section 5.4 to answer this question.
    b. For a sample of 80 taxpayers, what is the probability that at least 60 taxpayers say that it is very important to ensure that high-income tax payers do not cheat on their tax returns? Use the normal approximation of the binomial distribution to answer this question.
    c. As the number of trails in a binomial distribution application becomes large, what is the advantage of using the normal approximation of the binomial distribution to compute probabilities?
    d. When the number of trials for a binominal distribution application becomes large, would developers of statistical software packages prefer to use the binomial distribution probability function shown in Section 5.4 or the normal approximation of the binomial distribution shown in Section 6.3? Explain.
30. When you sign up for a credit card, do you read the contract carefully? In a FindLaw.com survey, individuals were asked, "How closely do you read a contract for a credit card?" (*USA Today,* October 16, 2003). The findings were that 44% read every word, 33% read enough to understand the contract, 11% just glance at it, and 4% don't read it at all.
    a. For a sample of 500 people, how many would you expect to say that they read every word of a credit card contract?
    b. For a sample of 500 people, what is the probability that 200 or fewer will say they read every word of a credit card contract?
    c. For a sample of 500 people, what is the probability that at least 15 say they don't read credit card contracts?
31. A Myrtle Beach resort hotel has 120 rooms. In the spring months, hotel room occupancy is approximately 75%.
    a. What is the probability that at least half of the rooms are occupied on a given day?
    b. What is the probability that 100 or more rooms are occupied on a given day?
    c. What is the probability that 80 or fewer rooms are occupied on a given day?

6.4 Exponential Probability Distribution

The **exponential probability distribution** may be used for random variables such as the time between arrivals at a car wash, the time required to load a truck, the distance between major defects in a highway, and so on. The exponential probability density function follows.

EXPONENTIAL PROBABILITY DENSITY FUNCTION

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \qquad \text{for } x \geq 0 \tag{6.4}$$

where μ = expected value or mean

As an example of the exponential distribution, suppose that x represents the loading time for a truck at the Schips loading dock and follows such a distribution. If the mean, or average, loading time is 15 minutes ($\mu = 15$), the appropriate probability density function for x is

$$f(x) = \frac{1}{15} e^{-x/15}$$

Figure 6.10 is the graph of this probability density function.

Computing Probabilities for the Exponential Distribution

In waiting line applications, the exponential distribution is often used for service time.

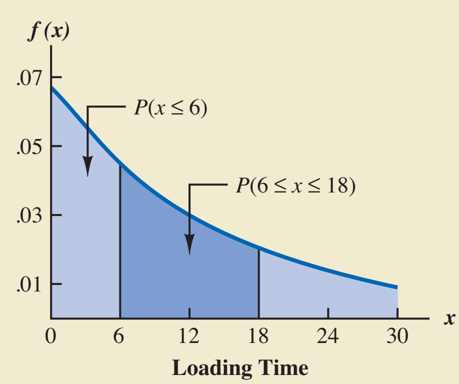
As with any continuous probability distribution, the area under the curve corresponding to an interval provides the probability that the random variable assumes a value in that interval. In the Schips loading dock example, the probability that loading a truck will take 6 minutes or less $P(x \leq 6)$ is defined to be the area under the curve in Figure 6.10 from $x = 0$ to $x = 6$. Similarly, the probability that the loading time will be 18 minutes or less $P(x \leq 18)$ is the area under the curve from $x = 0$ to $x = 18$. Note also that the probability that the loading time will be between 6 minutes and 18 minutes $P(6 \leq x \leq 18)$ is given by the area under the curve from $x = 6$ to $x = 18$.

To compute exponential probabilities such as those just described, we use the following formula. It provides the cumulative probability of obtaining a value for the exponential random variable of less than or equal to some specific value denoted by x_0.

EXPONENTIAL DISTRIBUTION: CUMULATIVE PROBABILITIES

$$P(x \leq x_0) = 1 - e^{-x_0/\mu} \tag{6.5}$$

FIGURE 6.10 EXPONENTIAL DISTRIBUTION FOR THE SCHIPS LOADING DOCK EXAMPLE



For the Schips loading dock example, x = loading time in minutes and μ = 15 minutes. Using equation (6.5)

$$P(x \leq x_0) = 1 - e^{-x_0/15}$$

Hence, the probability that loading a truck will take 6 minutes or less is

$$P(x \leq 6) = 1 - e^{-6/15} = .3297$$

Using equation (6.5), we calculate the probability of loading a truck in 18 minutes or less.

$$P(x \leq 18) = 1 - e^{-18/15} = .6988$$

Thus, the probability that loading a truck will take between 6 minutes and 18 minutes is equal to $.6988 - .3297 = .3691$. Probabilities for any other interval can be computed similarly.

A property of the exponential distribution is that the mean and standard deviation are equal.

In the preceding example, the mean time it takes to load a truck is $\mu = 15$ minutes. A property of the exponential distribution is that the mean of the distribution and the standard deviation of the distribution are *equal*. Thus, the standard deviation for the time it takes to load a truck is $\sigma = 15$ minutes. The variance is $\sigma^2 = (15)^2 = 225$.

Relationship Between the Poisson and Exponential Distributions

In Section 5.5 we introduced the Poisson distribution as a discrete probability distribution that is often useful in examining the number of occurrences of an event over a specified interval of time or space. Recall that the Poisson probability function is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}$$

where

μ = expected value or mean number of occurrences over a specified interval

If arrivals follow a Poisson distribution, the time between arrivals must follow an exponential distribution.

The continuous exponential probability distribution is related to the discrete Poisson distribution. If the Poisson distribution provides an appropriate description of the number of occurrences per interval, the exponential distribution provides a description of the length of the interval between occurrences.

To illustrate this relationship, suppose the number of cars that arrive at a car wash during one hour is described by a Poisson probability distribution with a mean of 10 cars per hour. The Poisson probability function that gives the probability of x arrivals per hour is

$$f(x) = \frac{10^x e^{-10}}{x!}$$

Because the average number of arrivals is 10 cars per hour, the average time between cars arriving is

$$\frac{1 \text{ hour}}{10 \text{ cars}} = .1 \text{ hour/car}$$

Thus, the corresponding exponential distribution that describes the time between the arrivals has a mean of $\mu = .1$ hour per car; as a result, the appropriate exponential probability density function is

$$f(x) = \frac{1}{.1} e^{-x/.1} = 10e^{-10x}$$

NOTES AND COMMENTS

As we can see in Figure 6.10, the exponential distribution is skewed to the right. Indeed, the skewness measure for exponential distributions is 2. The exponential distribution gives us a good idea what a skewed distribution looks like.

Exercises

Methods

32. Consider the following exponential probability density function.

$$f(x) = \frac{1}{8} e^{-x/8} \qquad \text{for } x \geq 0$$

a. Find $P(x \leq 6)$.
b. Find $P(x \leq 4)$.
c. Find $P(x \geq 6)$.
d. Find $P(4 \leq x \leq 6)$.

SELF test

33. Consider the following exponential probability density function.

$$f(x) = \frac{1}{3} e^{-x/3} \qquad \text{for } x \geq 0$$

a. Write the formula for $P(x \leq x_0)$.
b. Find $P(x \leq 2)$.
c. Find $P(x \geq 3)$.
d. Find $P(x \leq 5)$.
e. Find $P(2 \leq x \leq 5)$.

Applications

34. The time required to pass through security screening at the airport can be annoying to travelers. The mean wait time during peak periods at Cincinnati/Northern Kentucky International Airport is 12.1 minutes (*The Cincinnati Enquirer,* February 2, 2006). Assume the time to pass through security screening follows an exponential distribution.
    a. What is the probability it will take less than 10 minutes to pass through security screening during a peak period?
    b. What is the probability it will take more than 20 minutes to pass through security screening during a peak period?
    c. What is the probability it will take between 10 and 20 minutes to pass through security screening during a peak period?
    d. It is 8:00 A.M. (a peak period) and you just entered the security line. To catch your plane you must be at the gate within 30 minutes. If it takes 12 minutes from the time you clear security until you reach your gate, what is the probability you will miss your flight?

SELF test

35. The time between arrivals of vehicles at a particular intersection follows an exponential probability distribution with a mean of 12 seconds.
    a. Sketch this exponential probability distribution.
    b. What is the probability that the arrival time between vehicles is 12 seconds or less?
    c. What is the probability that the arrival time between vehicles is 6 seconds or less?
    d. What is the probability of 30 or more seconds between vehicle arrivals?

36. Comcast Corporation is the largest cable television company, the second largest Internet service provider, and the fourth largest telephone service provider in the United States. Generally known for quality and reliable service, the company periodically experiences unexpected service interruptions. On January 14, 2009, such an interruption occurred for the Comcast customers living in southwest Florida. When customers called the Comcast office, a recorded message told them that the company was aware of the service outage and that it was anticipated that service would be restored in two hours. Assume that two hours is the mean time to do the repair and that the repair time has an exponential probability distribution.
    a. What is the probability that the cable service will be repaired in one hour or less?
    b. What is the probability that the repair will take between one hour and two hours?
    c. For a customer who calls the Comcast office at 1:00 P.M., what is the probability that the cable service will not be repaired by 5:00 P.M.?
37. Collina's Italian Café in Houston, Texas, advertises that carryout orders take about 25 minutes (Collina's website, February 27, 2008). Assume that the time required for a carryout order to be ready for customer pickup has an exponential distribution with a mean of 25 minutes.
    a. What is the probability than a carryout order will be ready within 20 minutes?
    b. If a customer arrives 30 minutes after placing an order, what is the probability that the order will not be ready?
    c. A particular customer lives 15 minutes from Collina's Italian Café. If the customer places a telephone order at 5:20 P.M., what is the probability that the customer can drive to the café, pick up the order, and return home by 6:00 P.M.?
38. Do interruptions while you are working reduce your productivity? According to a University of California–Irvine study, businesspeople are interrupted at the rate of approximately 5½ times per hour (*Fortune,* March 20, 2006). Suppose the number of interruptions follows a Poisson probability distribution.
    a. Show the probability distribution for the time between interruptions.
    b. What is the probability a businessperson will have no interruptions during a 15-minute period?
    c. What is the probability that the next interruption will occur within 10 minutes for a particular businessperson?

Summary

This chapter extended the discussion of probability distributions to the case of continuous random variables. The major conceptual difference between discrete and continuous probability distributions involves the method of computing probabilities. With discrete distributions, the probability function $f(x)$ provides the probability that the random variable x assumes various values. With continuous distributions, the probability density function $f(x)$ does not provide probability values directly. Instead, probabilities are given by areas under the curve or graph of the probability density function $f(x)$. Because the area under the curve above a single point is zero, we observe that the probability of any particular value is zero for a continuous random variable.

Three continuous probability distributions—the uniform, normal, and exponential distributions—were treated in detail. The normal distribution is used widely in statistical inference and will be used extensively throughout the remainder of the text.

Glossary

Probability density function A function used to compute probabilities for a continuous random variable. The area under the graph of a probability density function over an interval represents probability.

Uniform probability distribution A continuous probability distribution for which the probability that the random variable will assume a value in any interval is the same for each interval of equal length.

Normal probability distribution A continuous probability distribution. Its probability density function is bell-shaped and determined by its mean μ and standard deviation σ.

Standard normal probability distribution A normal distribution with a mean of zero and a standard deviation of one.

Continuity correction factor A value of .5 that is added to or subtracted from a value of x when the continuous normal distribution is used to approximate the discrete binomial distribution.

Exponential probability distribution A continuous probability distribution that is useful in computing probabilities for the time it takes to complete a task.

Key Formulas

Uniform Probability Density Function

$$f(x) = \begin{cases} \dfrac{1}{b - a} & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases} \tag{6.1}$$

Normal Probability Density Function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{6.2}$$

Converting to the Standard Normal Random Variable

$$z = \frac{x - \mu}{\sigma} \tag{6.3}$$

Exponential Probability Density Function

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \qquad \text{for } x \geq 0 \tag{6.4}$$

Exponential Distribution: Cumulative Probabilities

$$P(x \leq x_0) = 1 - e^{-x_0/\mu} \tag{6.5}$$

Supplementary Exercises

39. A business executive, transferred from Chicago to Atlanta, needs to sell her house in Chicago quickly. The executive's employer has offered to buy the house for $210,000, but the offer expires at the end of the week. The executive does not currently have a better offer but can afford to leave the house on the market for another month. From conversations with

her realtor, the executive believes the price she will get by leaving the house on the market for another month is uniformly distributed between \$200,000 and \$225,000.

a. If she leaves the house on the market for another month, what is the mathematical expression for the probability density function of the sales price?
b. If she leaves it on the market for another month, what is the probability she will get at least \$215,000 for the house?
c. If she leaves it on the market for another month, what is the probability she will get less than \$210,000?
d. Should the executive leave the house on the market for another month? Why or why not?

40. The U.S. Bureau of Labor Statistics reports that the average annual expenditure on food and drink for all families is \$5700 (*Money,* December 2003). Assume that annual expenditure on food and drink is normally distributed and that the standard deviation is \$1500.
    a. What is the range of expenditures of the 10% of families with the lowest annual spending on food and drink?
    b. What percentage of families spend more than \$7000 annually on food and drink?
    c. What is the range of expenditures for the 5% of families with the highest annual spending on food and drink?

41. Motorola used the normal distribution to determine the probability of defects and the number of defects expected in a production process. Assume a production process produces items with a mean weight of 10 ounces. Calculate the probability of a defect and the expected number of defects for a 1000-unit production run in the following situations.
    a. The process standard deviation is .15, and the process control is set at plus or minus one standard deviation. Units with weights less than 9.85 or greater than 10.15 ounces will be classified as defects.
    b. Through process design improvements, the process standard deviation can be reduced to .05. Assume the process control remains the same, with weights less than 9.85 or greater than 10.15 ounces being classified as defects.
    c. What is the advantage of reducing process variation, thereby causing process control limits to be at a greater number of standard deviations from the mean?

42. The average annual amount American households spend for daily transportation is \$6312 (*Money,* August 2001). Assume that the amount spent is normally distributed.
    a. Suppose you learn that 5% of American households spend less than \$1000 for daily transportation. What is the standard deviation of the amount spent?
    b. What is the probability that a household spends between \$4000 and \$6000?
    c. What is the range of spending for the 3% of households with the highest daily transportation cost?

43. *Condé Nast Traveler* publishes a Gold List of the top hotels all over the world. The Broadmoor Hotel in Colorado Springs contains 700 rooms and is on the 2004 Gold List (*Condé Nast Traveler,* January 2004). Suppose Broadmoor's marketing group forecasts a mean demand of 670 rooms for the coming weekend. Assume that demand for the upcoming weekend is normally distributed with a standard deviation of 30.
    a. What is the probability all the hotel's rooms will be rented?
    b. What is the probability 50 or more rooms will not be rented?
    c. Would you recommend the hotel consider offering a promotion to increase demand? What considerations would be important?

44. Ward Doering Auto Sales is considering offering a special service contract that will cover the total cost of any service work required on leased vehicles. From experience, the company manager estimates that yearly service costs are approximately normally distributed, with a mean of \$150 and a standard deviation of \$25.
    a. If the company offers the service contract to customers for a yearly charge of \$200, what is the probability that any one customer's service costs will exceed the contract price of \$200?
    b. What is Ward's expected profit per service contract?

45. Is lack of sleep causing traffic fatalities? A study conducted under the auspices of the National Highway Traffic Safety Administration found that the average number of fatal crashes caused by drowsy drivers each year was 1550 (*BusinessWeek,* January 26, 2004). Assume the annual number of fatal crashes per year is normally distributed with a standard deviation of 300.
    a. What is the probability of fewer than 1000 fatal crashes in a year?
    b. What is the probability the number of fatal crashes will be between 1000 and 2000 for a year?
    c. For a year to be in the upper 5% with respect to the number of fatal crashes, how many fatal crashes would have to occur?
46. Assume that the test scores from a college admissions test are normally distributed, with a mean of 450 and a standard deviation of 100.
    a. What percentage of the people taking the test score between 400 and 500?
    b. Suppose someone receives a score of 630. What percentage of the people taking the test score better? What percentage score worse?
    c. If a particular university will not admit anyone scoring below 480, what percentage of the persons taking the test would be acceptable to the university?
47. According to Salary Wizard, the average base salary for a brand manager in Houston, Texas, is $88,592 and the average base salary for a brand manager in Los Angeles, California, is $97,417 (Salary Wizard website*,* February 27, 2008). Assume that salaries are normally distributed, the standard deviation for brand managers in Houston is $19,900, and the standard deviation for brand managers in Los Angeles is $21,800.
    a. What is the probability that a brand manager in Houston has a base salary in excess of $100,000?
    b. What is the probability that a brand manager in Los Angeles has a base salary in excess of $100,000?
    c. What is the probability that a brand manager in Los Angeles has a base salary of less than $75,000?
    d. How much would a brand manager in Los Angeles have to make in order to have a higher salary than 99% of the brand managers in Houston?
48. A machine fills containers with a particular product. The standard deviation of filling weights is known from past data to be .6 ounce. If only 2% of the containers hold less than 18 ounces, what is the mean filling weight for the machine? That is, what must μ equal? Assume the filling weights have a normal distribution.
49. Consider a multiple-choice examination with 50 questions. Each question has four possible answers. Assume that a student who has done the homework and attended lectures has a 75% probability of answering any question correctly.
    a. A student must answer 43 or more questions correctly to obtain a grade of A. What percentage of the students who have done their homework and attended lectures will obtain a grade of A on this multiple-choice examination?
    b. A student who answers 35 to 39 questions correctly will receive a grade of C. What percentage of students who have done their homework and attended lectures will obtain a grade of C on this multiple-choice examination?
    c. A student must answer 30 or more questions correctly to pass the examination. What percentage of the students who have done their homework and attended lectures will pass the examination?
    d. Assume that a student has not attended class and has not done the homework for the course. Furthermore, assume that the student will simply guess at the answer to each question. What is the probability that this student will answer 30 or more questions correctly and pass the examination?
50. A blackjack player at a Las Vegas casino learned that the house will provide a free room if play is for four hours at an average bet of $50. The player's strategy provides a

probability of .49 of winning on any one hand, and the player knows that there are 60 hands per hour. Suppose the player plays for four hours at a bet of \$50 per hand.

a. What is the player's expected payoff?
b. What is the probability the player loses \$1000 or more?
c. What is the probability the player wins?
d. Suppose the player starts with \$1500. What is the probability of going broke?

51. The time in minutes for which a student uses a computer terminal at the computer center of a major university follows an exponential probability distribution with a mean of 36 minutes. Assume a student arrives at the terminal just as another student is beginning to work on the terminal.
    a. What is the probability that the wait for the second student will be 15 minutes or less?
    b. What is the probability that the wait for the second student will be between 15 and 45 minutes?
    c. What is the probability that the second student will have to wait an hour or more?

52. The website for the Bed and Breakfast Inns of North America gets approximately seven visitors per minute (*Time,* September 2001). Suppose the number of website visitors per minute follows a Poisson probability distribution.
    a. What is the mean time between visits to the website?
    b. Show the exponential probability density function for the time between website visits.
    c. What is the probability no one will access the website in a 1-minute period?
    d. What is the probability no one will access the website in a 12-second period?

53. The American Community Survey showed that residents of New York City have the longest travel times to get to work compared to residents of other cities in the United States (U.S. Census Bureau website, August 2008). According to the latest statistics available, the average travel time to work for residents of New York City is 38.3 minutes.
    a. Assume the exponential probability distribution is applicable and show the probability density function for the travel time to work for a resident of this city.
    b. What is the probability it will take a resident of this city between 20 and 40 minutes to travel to work?
    c. What is the probability it will take a resident of this city more than one hour to travel to work?

54. The time (in minutes) between telephone calls at an insurance claims office has the following exponential probability distribution.

$$f(x) = .50e^{-.50x} \qquad \text{for } x \geq 0$$

    a. What is the mean time between telephone calls?
    b. What is the probability of having 30 seconds or less between telephone calls?
    c. What is the probability of having 1 minute or less between telephone calls?
    d. What is the probability of having 5 or more minutes without a telephone call?

Case Problem Specialty Toys

Specialty Toys, Inc., sells a variety of new and innovative children's toys. Management learned that the preholiday season is the best time to introduce a new toy, because many families use this time to look for new ideas for December holiday gifts. When Specialty discovers a new toy with good market potential, it chooses an October market entry date.

In order to get toys in its stores by October, Specialty places one-time orders with its manufacturers in June or July of each year. Demand for children's toys can be highly volatile. If a new toy catches on, a sense of shortage in the marketplace often increases the demand to high levels and large profits can be realized. However, new toys can also flop, leaving Specialty stuck with high levels of inventory that must be sold at reduced prices. The most important question the company faces is deciding how many units of a new toy should be purchased to meet anticipated sales demand. If too few are purchased, sales will be lost; if too many are purchased, profits will be reduced because of low prices realized in clearance sales.

For the coming season, Specialty plans to introduce a new product called Weather Teddy. This variation of a talking teddy bear is made by a company in Taiwan. When a child presses Teddy's hand, the bear begins to talk. A built-in barometer selects one of five responses that predict the weather conditions. The responses range from "It looks to be a very nice day! Have fun" to "I think it may rain today. Don't forget your umbrella." Tests with the product show that, even though it is not a perfect weather predictor, its predictions are surprisingly good. Several of Specialty's managers claimed Teddy gave predictions of the weather that were as good as many local television weather forecasters.

As with other products, Specialty faces the decision of how many Weather Teddy units to order for the coming holiday season. Members of the management team suggested order quantities of 15,000, 18,000, 24,000, or 28,000 units. The wide range of order quantities suggested indicates considerable disagreement concerning the market potential. The product management team asks you for an analysis of the stock-out probabilities for various order quantities, an estimate of the profit potential, and to help make an order quantity recommendation. Specialty expects to sell Weather Teddy for \$24 based on a cost of \$16 per unit. If inventory remains after the holiday season, Specialty will sell all surplus inventory for \$5 per unit. After reviewing the sales history of similar products, Specialty's senior sales forecaster predicted an expected demand of 20,000 units with a .95 probability that demand would be between 10,000 units and 30,000 units.

Managerial Report

Prepare a managerial report that addresses the following issues and recommends an order quantity for the Weather Teddy product.

1. Use the sales forecaster's prediction to describe a normal probability distribution that can be used to approximate the demand distribution. Sketch the distribution and show its mean and standard deviation.
2. Compute the probability of a stock-out for the order quantities suggested by members of the management team.
3. Compute the projected profit for the order quantities suggested by the management team under three scenarios: worst case in which sales = 10,000 units, most likely case in which sales = 20,000 units, and best case in which sales = 30,000 units.
4. One of Specialty's managers felt that the profit potential was so great that the order quantity should have a 70% chance of meeting demand and only a 30% chance of any stock-outs. What quantity would be ordered under this policy, and what is the projected profit under the three sales scenarios?
5. Provide your own recommendation for an order quantity and note the associated profit projections. Provide a rationale for your recommendation.

Appendix 6.1 Continuous Probability Distributions with Minitab

Let us demonstrate the Minitab procedure for computing continuous probabilities by referring to the Grear Tire Company problem where tire mileage was described by a normal distribution with $\mu = 36{,}500$ and $\sigma = 5000$. One question asked was: What is the probability that the tire mileage will exceed 40,000 miles?

For continuous probability distributions, Minitab gives a cumulative probability; that is, Minitab gives the probability that the random variable will assume a value less than or equal to a specified constant. For the Grear tire mileage question, Minitab can be used to determine the cumulative probability that the tire mileage will be less than or equal to 40,000 miles. (The specified constant in this case is 40,000.) After obtaining the cumulative probability from Minitab, we must subtract it from 1 to determine the probability that the tire mileage will exceed 40,000 miles.

Prior to using Minitab to compute a probability, one must enter the specified constant into a column of the worksheet. For the Grear tire mileage question we entered the specified constant of 40,000 into column C1 of the Minitab worksheet. The steps in using Minitab to compute the cumulative probability of the normal random variable assuming a value less than or equal to 40,000 follow.

Step 1. Select the **Calc** menu
Step 2. Choose **Probability Distributions**
Step 3. Choose **Normal**
Step 4. When the Normal Distribution dialog box appears:
Select **Cumulative probability**
Enter 36500 in the **Mean** box
Enter 5000 in the **Standard deviation** box
Enter C1 in the **Input column** box (the column containing 40,000)
Click **OK**

After the user clicks **OK,** Minitab prints the cumulative probability that the normal random variable assumes a value less than or equal to 40,000. Minitab shows that this probability is .7580. Because we are interested in the probability that the tire mileage will be greater than 40,000, the desired probability is $1 - .7580 = .2420$.

A second question in the Grear Tire Company problem was: What mileage guarantee should Grear set to ensure that no more than 10% of the tires qualify for the guarantee? Here we are given a probability and want to find the corresponding value for the random variable. Minitab uses an inverse calculation routine to find the value of the random variable associated with a given cumulative probability. First, we must enter the cumulative probability into a column of the Minitab worksheet (say, C1). In this case, the desired cumulative probability is .10. Then, the first three steps of the Minitab procedure are as already listed. In step 4, we select **Inverse cumulative probability** instead of **Cumulative probability** and complete the remaining parts of the step. Minitab then displays the mileage guarantee of 30,092 miles.

Minitab is capable of computing probabilities for other continuous probability distributions, including the exponential probability distribution. To compute exponential probabilities, follow the procedure shown previously for the normal probability distribution and choose the **Exponential** option in step 3. Step 4 is as shown, with the exception that entering the standard deviation is not required. Output for cumulative probabilities and inverse cumulative probabilities is identical to that described for the normal probability distribution.

Appendix 6.2 Continuous Probability Distributions with Excel

Excel provides the capability for computing probabilities for several continuous probability distributions, including the normal and exponential probability distributions. In this appendix, we describe how Excel can be used to compute probabilities for any normal distribution. The procedures for the exponential and other continuous distributions are similar to the one we describe for the normal distribution.

Let us return to the Grear Tire Company problem where the tire mileage was described by a normal distribution with $\mu = 36{,}500$ and $\sigma = 5000$. Assume we are interested in the probability that tire mileage will exceed 40,000 miles.

Excel's NORMDIST function provides cumulative probabilities for a normal distribution. The general form of the function is NORMDIST (x,μ,σ,cumulative). For the fourth argument, TRUE is specified if a cumulative probability is desired. Thus, to compute the cumulative probability that the tire mileage will be less than or equal to 40,000 miles we would enter the following formula into any cell of an Excel worksheet:

=NORMDIST(40000,36500,5000,TRUE)

At this point, .7580 will appear in the cell where the formula was entered, indicating that the probability of tire mileage being less than or equal to 40,000 miles is .7580. Therefore, the probability that tire mileage will exceed 40,000 miles is $1 - .7580 = .2420$.

Excel's NORMINV function uses an inverse computation to find the x value corresponding to a given cumulative probability. For instance, suppose we want to find the guaranteed mileage Grear should offer so that no more than 10% of the tires will be eligible for the guarantee. We would enter the following formula into any cell of an Excel worksheet:

=NORMINV(.1,36500,5000)

At this point, 30092 will appear in the cell where the formula was entered, indicating that the probability of a tire lasting 30,092 miles or less is .10.

The Excel function for computing exponential probabilities is EXPONDIST. Using it is straightforward. But if one needs help specifying the proper values for the arguments, Excel's Insert Function dialog box can be used (see Appendix E).