CHAPTER 6

Continuous Distributions

LEARNING OBJECTIVES

The primary learning objective of Chapter 6 is to help you understand continuous distributions, thereby enabling you to:

1. Solve for probabilities in a continuous uniform distribution
2. Solve for probabilities in a normal distribution using *z* scores and for the mean, the standard deviation, or a value of *x* in a normal distribution when given information about the area under the normal curve
3. Solve problems from the discrete binomial distribution using the continuous normal distribution and correcting for continuity
4. Solve for probabilities in an exponential distribution and contrast the exponential distribution to the discrete Poisson distribution

Stockbyte/Getty Images

The Cost of Human Resources

Decision Dilemma

What is the human resource cost of hiring and maintaining employees in a company? Studies conducted by Saratoga Institute, Pricewaterhouse-Coopers Human Resource Services, determined that the average cost of hiring an employee is $3,270, and the average annual human resource expenditure per employee is $1,554. The average health benefit payment per employee is $6,393, and the average employer 401(k) cost per participant is $2,258. According to a survey conducted by the American Society for Training and Development, companies annually spend an average of $955 per employee on training, and, on average, an employee receives 32 hours of training annually. Business researchers have attempted to measure the cost of employee absenteeism to an organization. A survey conducted by CCH, Inc., showed that the average annual cost of unscheduled absenteeism per employee is $660. According to this survey, 35% of all unscheduled absenteeism is caused by personal illness.

Managerial and Statistical Questions

1. The survey conducted by the American Society for Training and Development reported that, on average, an employee receives 32 hours of training per year. Suppose that number of hours of training is uniformly distributed across all employees varying from 0 hours to 64 hours. What percentage of employees receive between 20 and 40 hours of training? What percentage of employees receive 50 hours or more of training?
2. As the result of another survey, it was estimated that, on average, it costs $3,270 to hire an employee. Suppose such costs are normally distributed with a standard deviation of $400. Based on these figures, what is the probability that a randomly selected hire costs more than $4,000? What percentage of employees is hired for less than $3,000?
3. The absenteeism survey determined that 35% of all unscheduled absenteeism is caused by personal illness. If this is true, what is the probability of randomly sampling 120 unscheduled absences and finding out that more than 50 were caused by personal illness?

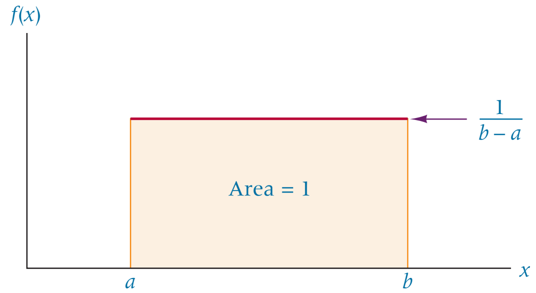
Sources: Adapted from: "Human Resources Is Put on Notice," *Workforce Management,* vol. 84, no. 14 (December 12, 2005), pp. 44–48. Web sites for data sources: www.pwcservices.com/, www.astd.org, and www.cch.com

Whereas Chapter 5 focused on the characteristics and applications of discrete distributions, Chapter 6 concentrates on information about continuous distributions. Continuous distributions are constructed from continuous random variables in which values are taken on for every point over a given interval and are usually generated from experiments in which things are "measured" as opposed to "counted." With continuous distributions, probabilities of outcomes occurring between particular points are determined by calculating the area under the curve between those points. In addition, the entire area under the whole curve is equal to 1. The many continuous distributions in statistics include the uniform distribution, the normal distribution, the exponential distribution, the t distribution, the chi-square distribution, and the F distribution. This chapter presents the uniform distribution, the normal distribution, and the exponential distribution.

6.1 THE UNIFORM DISTRIBUTION

The **uniform distribution**, sometimes referred to as the **rectangular distribution**, is *a relatively simple continuous distribution in which the same height, or $f(x)$, is obtained over a range of values.* The following probability density function defines a uniform distribution.

FIGURE 6.1
Uniform Distribution



PROBABILITY DENSITY FUNCTION OF A UNIFORM DISTRIBUTION

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for all other values} \end{cases}$$

Figure 6.1 is an example of a uniform distribution. In a uniform, or rectangular, distribution, the total area under the curve is equal to the product of the length and the width of the rectangle and equals 1. Because the distribution lies, by definition, between the x values of a and b, the length of the rectangle is $(b - a)$. Combining this area calculation with the fact that the area equals 1, the height of the rectangle can be solved as follows.

$$\text{Area of Rectangle} = (\text{Length})(\text{Height}) = 1$$

But

$$\text{Length} = (b - a)$$

Therefore,

$$(b - a)(\text{Height}) = 1$$

and

$$\text{Height} = \frac{1}{(b - a)}$$

These calculations show why, between the x values of a and b, the distribution has a constant height of $1/(b - a)$.

The mean and standard deviation of a uniform distribution are given as follows.

MEAN AND STANDARD DEVIATION OF A UNIFORM DISTRIBUTION

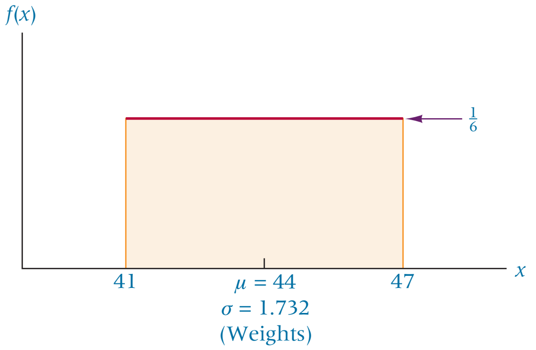
$$\mu = \frac{a + b}{2}$$

$$\sigma = \frac{b - a}{\sqrt{12}}$$

Many possible situations arise in which data might be uniformly distributed. As an example, suppose a production line is set up to manufacture machine braces in lots of five per minute during a shift. When the lots are weighed, variation among the weights is detected, with lot weights ranging from 41 to 47 grams in a uniform distribution. The height of this distribution is

$$f(x) = \text{Height} = \frac{1}{(b - a)} = \frac{1}{(47 - 41)} = \frac{1}{6}$$

FIGURE 6.2
Distribution of Lot Weights



The mean and standard deviation of this distribution are

$$\text{Mean} = \frac{a+b}{2} = \frac{41+47}{2} = \frac{88}{2} = 44$$

$$\text{Standard Deviation} = \frac{b-a}{\sqrt{12}} = \frac{47-41}{\sqrt{12}} = \frac{6}{3.464} = 1.732$$

Figure 6.2 provides the uniform distribution for this example, with its mean, standard deviation, and the height of the distribution.

Determining Probabilities in a Uniform Distribution

With discrete distributions, the probability function yields the value of the probability. For continuous distributions, probabilities are calculated by determining the area over an interval of the function. With continuous distributions, there is no area under the curve for a single point. The following equation is used to determine the probabilities of x for a uniform distribution between a and b.

PROBABILITIES IN A UNIFORM DISTRIBUTION

$$P(x) = \frac{x_2 - x_1}{b - a}$$

where:

$$a \leq x_1 \leq x_2 \leq b$$

Remember that the area between a and b is equal to 1. The probability for any interval that includes a and b is 1. The probability of $x \geq b$ or of $x \leq a$ is zero because there is no area above b or below a.

Suppose that on the machine braces problem we want to determine the probability that a lot weighs between 42 and 45 grams. This probability is computed as follows:

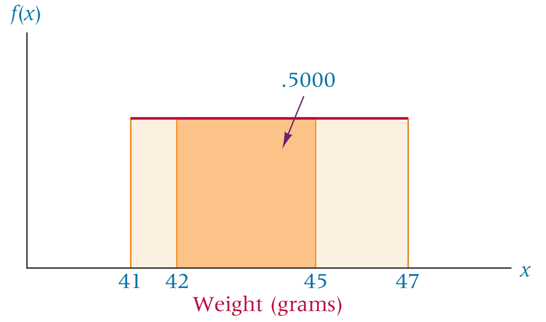
$$P(x) = \frac{x_2 - x_1}{b - a} = \frac{45 - 42}{47 - 41} = \frac{3}{6} = .5000$$

Figure 6.3 displays this solution.

The probability that a lot weighs more than 48 grams is zero, because $x = 48$ is greater than the upper value, $x = 47$, of the uniform distribution. A similar argument gives the probability of a lot weighing less than 40 grams. Because 40 is less than the lowest value of the uniform distribution range, 41, the probability is zero.

FIGURE 6.3

Solved Probability in a Uniform Distribution



DEMONSTRATION PROBLEM 6.1

Suppose the amount of time it takes to assemble a plastic module ranges from 27 to 39 seconds and that assembly times are uniformly distributed. Describe the distribution. What is the probability that a given assembly will take between 30 and 35 seconds? Fewer than 30 seconds?

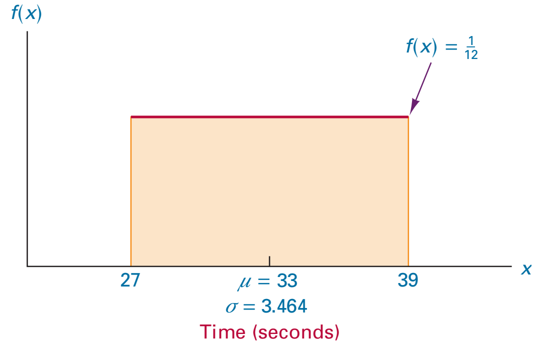
Solution

$$f(x) = \frac{1}{39 - 27} = \frac{1}{12}$$

$$\mu = \frac{a + b}{2} = \frac{39 + 27}{2} = 33$$

$$\sigma = \frac{b - a}{\sqrt{12}} = \frac{39 - 27}{\sqrt{12}} = \frac{12}{\sqrt{12}} = 3.464$$

The height of the distribution is 1/12. The mean time is 33 seconds with a standard deviation of 3.464 seconds.



$$P(30 \le x \le 35) = \frac{35 - 30}{39 - 27} = \frac{5}{12} = .4167$$

There is a .4167 probability that it will take between 30 and 35 seconds to assemble the module.

$$P(x < 30) = \frac{30 - 27}{39 - 27} = \frac{3}{12} = .2500$$

There is a .2500 probability that it will take less than 30 seconds to assemble the module. Because there is no area less than 27 seconds, $P(x < 30)$ is determined by using only the interval $27 \le x < 30$. In a continuous distribution, there is no area at any one point (only over an interval). Thus the probability $x < 30$ is the same as the probability of $x \le 30$.

DEMONSTRATION PROBLEM 6.2

According to the National Association of Insurance Commissioners, the average annual cost for automobile insurance in the United States in a recent year was $691. Suppose automobile insurance costs are uniformly distributed in the United States with a range of from $200 to $1,182. What is the standard deviation of this uniform distribution? What is the height of the distribution? What is the probability that a person's annual cost for automobile insurance in the United States is between $410 and $825?

Solution

The mean is given as $691. The value of a is $200 and b is $1,182.

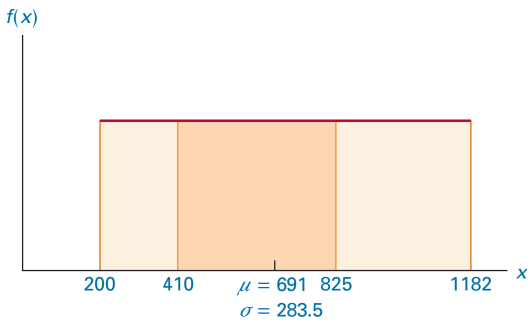
$$\sigma = \frac{b - a}{\sqrt{12}} = \frac{1{,}182 - 200}{\sqrt{12}} = 283.5$$

The height of the distribution is $\frac{1}{1{,}182 - 200} = \frac{1}{982} = .001$

$$x_1 = 410 \text{ and } x_2 = 825$$

$$P(410 \leq x \leq 825) = \frac{825 - 410}{1{,}182 - 200} = \frac{415}{982} = .4226$$

The probability that a randomly selected person pays between $410 and $825 annually for automobile insurance in the United States is .4226. That is, about 42.26% of all people in the United States pay in that range.



Using the Computer to Solve for Uniform Distribution Probabilities

Using the values of a, b, and x, Minitab has the capability of computing probabilities for the uniform distribution. The resulting computation is a cumulative probability from the left end of the distribution to each x value. As an example, the probability question, $P(410 \leq x \leq 825)$, from Demonstration Problem 6.2 can be worked using Minitab. Minitab computes the probability of $x \leq 825$ and the probability of $x \leq 410$, and these results are shown in Table 6.1. The final answer to the probability question from Demonstration Problem 6.2 is obtained by subtracting these two probabilities:

$$P(410 \leq x \leq 825) = .6365 - .2138 = .4227$$

TABLE 6.1
Minitab Output for Uniform Distribution

CUMULATIVE DISTRIBUTION FUNCTION

Continuous uniform on 200 to 1182

x	P(X <= x)
825	0.636456
410	0.213849

Excel does not have the capability of directly computing probabilities for the uniform distribution.

6.1 PROBLEMS

6.1 Values are uniformly distributed between 200 and 240.

a. What is the value of $f(x)$ for this distribution?

b. Determine the mean and standard deviation of this distribution.

c. Probability of $(x > 230) = ?$

d. Probability of $(205 \leq x \leq 220) = ?$

e. Probability of $(x \leq 225) = ?$

6.2 x is uniformly distributed over a range of values from 8 to 21.

a. What is the value of $f(x)$ for this distribution?

b. Determine the mean and standard deviation of this distribution.

c. Probability of $(10 \leq x < 17) = ?$

d. Probability of $(x < 22) = ?$

e. Probability of $(x \geq 7) = ?$

6.3 The retail price of a medium-sized box of a well-known brand of cornflakes ranges from \$2.80 to \$3.14. Assume these prices are uniformly distributed. What are the average price and standard deviation of prices in this distribution? If a price is randomly selected from this list, what is the probability that it will be between \$3.00 and \$3.10?

6.4 The average fill volume of a regular can of soft drink is 12 ounces. Suppose the fill volume of these cans ranges from 11.97 to 12.03 ounces and is uniformly distributed. What is the height of this distribution? What is the probability that a randomly selected can contains more than 12.01 ounces of fluid? What is the probability that the fill volume is between 11.98 and 12.01 ounces?

6.5 Suppose the average U.S. household spends \$2,100 a year on all types of insurance. Suppose the figures are uniformly distributed between the values of \$400 and \$3,800. What are the standard deviation and the height of this distribution? What proportion of households spends more than \$3,000 a year on insurance? More than \$4,000? Between \$700 and \$1,500?

6.2 NORMAL DISTRIBUTION

Video

Interactive Applet

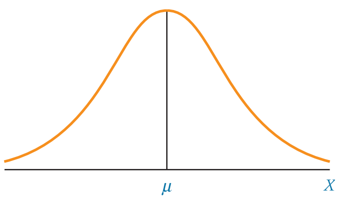
Probably the most widely known and used of all distributions is the **normal distribution**. It fits many human characteristics, such as height, weight, length, speed, IQ, scholastic achievement, and years of life expectancy, among others. Like their human counterparts, living things in nature, such as trees, animals, insects, and others, have many characteristics that are normally distributed.

Many variables in business and industry also are normally distributed. Some examples of variables that could produce normally distributed measurements include the annual cost of household insurance, the cost per square foot of renting warehouse space, and managers' satisfaction with support from ownership on a five-point scale. In addition, most items produced or filled by machines are normally distributed.

Because of its many applications, the normal distribution is an extremely important distribution. Besides the many variables mentioned that are normally distributed, the normal distribution and its associated probabilities are an integral part of statistical process control (see Chapter 18). When large enough sample sizes are taken, many statistics are normally distributed regardless of the shape of the underlying distribution from which they are drawn (as discussed in Chapter 7). Figure 6.4 is the graphic representation of the normal distribution: the normal curve.

FIGURE 6.4

The Normal Curve



History of the Normal Distribution

Discovery of the normal curve of errors is generally credited to mathematician and astronomer Karl Gauss (1777–1855), who recognized that the errors of repeated measurement of objects are often normally distributed.* Thus the normal distribution is sometimes referred to as the *Gaussian distribution* or the *normal curve of error*. A modern-day analogy of Gauss's work might be the distribution of measurements of machine-produced parts, which often yield a normal curve of error around a mean specification.

To a lesser extent, some credit has been given to Pierre-Simon de Laplace (1749–1827) for discovering the normal distribution. However, many people now believe that Abraham de Moivre (1667–1754), a French mathematician, first understood the normal distribution. De Moivre determined that the binomial distribution approached the normal distribution as a limit. De Moivre worked with remarkable accuracy. His published table values for the normal curve are only a few ten-thousandths off the values of currently published tables.†

The normal distribution exhibits the following characteristics.

- It is a continuous distribution.
- It is a symmetrical distribution about its mean.
- It is asymptotic to the horizontal axis.
- It is unimodal.
- It is a family of curves.
- Area under the curve is 1.

The normal distribution is symmetrical. Each half of the distribution is a mirror image of the other half. Many normal distribution tables contain probability values for only one side of the distribution because probability values for the other side of the distribution are identical because of symmetry.

In theory, the normal distribution is asymptotic to the horizontal axis. That is, it does not touch the x-axis, and it goes forever in each direction. The reality is that most applications of the normal curve are experiments that have finite limits of potential outcomes. For example, even though GMAT scores are analyzed by the normal distribution, the range of scores on each part of the GMAT is from 200 to 800.

The normal curve sometimes is referred to as the *bell-shaped curve*. It is unimodal in that values *mound up* in only one portion of the graph—the center of the curve. The normal distribution actually is a family of curves. Every unique value of the mean and every unique value of the standard deviation result in a different normal curve. In addition, *the total area under any normal distribution is 1*. The area under the curve yields the probabilities, so the total of all probabilities for a normal distribution is 1. Because the distribution is symmetric, the area of the distribution on each side of the mean is 0.5.

Probability Density Function of the Normal Distribution

The normal distribution is described or characterized by two parameters: the mean, μ, and the standard deviation, σ. The values of μ and σ produce a normal distribution. The density function of the normal distribution is

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

where

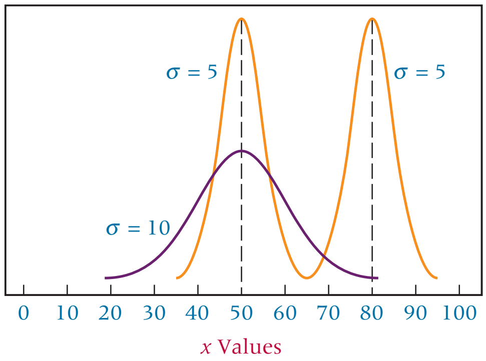
μ = mean of x
σ = standard deviation of x
π = 3.14159 . . . , and
e = 2.71828. . . .

*John A. Ingram and Joseph G. Monks, *Statistics for Business and Economics*. San Diego: Harcourt Brace Jovanovich, 1989.

†Roger E. Kirk, *Statistical Issues: A Reader for the Behavioral Sciences*. Monterey, CA: Brooks/Cole, 1972.

FIGURE 6.5

Normal Curves for Three Different Combinations of Means and Standard Deviations



Using Integral Calculus to determine areas under the normal curve from this function is difficult and time-consuming, therefore, virtually all researchers use table values to analyze normal distribution problems rather than this formula.

Standardized Normal Distribution

Every unique pair of μ and σ values defines a different normal distribution. Figure 6.5 shows the Minitab graphs of normal distributions for the following three pairs of parameters.

1. $\mu = 50$ and $\sigma = 5$
2. $\mu = 80$ and $\sigma = 5$
3. $\mu = 50$ and $\sigma = 10$

Note that every change in a parameter (μ or σ) determines a different normal distribution. This characteristic of the normal curve (a family of curves) could make analysis by the normal distribution tedious because volumes of normal curve tables—one for each different combination of μ and σ—would be required. Fortunately, a mechanism was developed by which all normal distributions can be converted into a single distribution: the z distribution. This process yields the **standardized normal distribution** (or curve). The conversion formula for any x value of a given normal distribution follows.

***z* FORMULA**

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

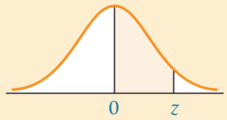
A ***z* score** is *the number of standard deviations that a value, x, is above or below the mean.* If the value of x is less than the mean, the z score is negative; if the value of x is more than the mean, the z score is positive; and if the value of x equals the mean, the associated z score is zero. This formula allows conversion of the distance of any x value from its mean into standard deviation units. A standard z score table can be used to find probabilities for any normal curve problem that has been converted to z scores. The ***z* distribution** is *a normal distribution with a mean of 0 and a standard deviation of 1.* Any value of x at the mean of a normal curve is zero standard deviations from the mean. Any value of x that is one standard deviation above the mean has a z value of 1. The empirical rule, introduced in Chapter 3, is based on the normal distribution in which about 68% of all values are within one standard deviation of the mean regardless of the values of μ and σ. In a z distribution, about 68% of the z values are between $z = -1$ and $z = +1$.

The z distribution probability values are given in Table A.5. Because it is so frequently used, the z distribution is also printed inside the cover of this text. For discussion purposes, a list of z distribution values is presented in Table 6.2.

Table A.5 gives the total area under the z curve between 0 and any point on the positive z-axis. Since the curve is symmetric, the area under the curve between z and 0 is the same whether z is positive or negative (the sign on the z value designates whether the z score is above or below the mean). The table areas or probabilities are always positive.

TABLE 6.2

z Distribution

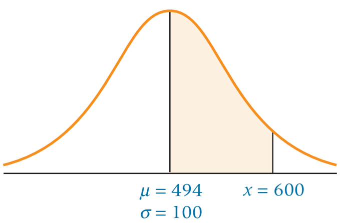


SECOND DECIMAL PLACE IN *z*

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

FIGURE 6.6

Graphical Depiction of the Area Between a Score of 600 and a Mean on a GMAT



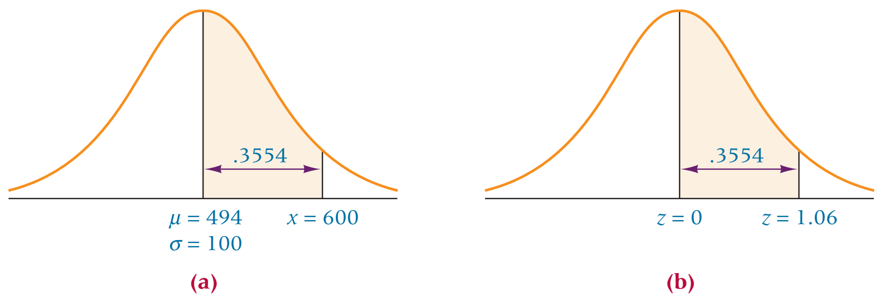
Solving Normal Curve Problems

The mean and standard deviation of a normal distribution and the *z* formula and table enable a researcher to determine the probabilities for intervals of any particular values of a normal curve. One example is the many possible probability values of GMAT scores examined next.

The Graduate Management Aptitude Test (GMAT), produced by the Educational Testing Service in Princeton, New Jersey, is widely used by graduate schools of business in

FIGURE 6.7

Graphical Solutions to the GMAT Problem



the United States as an entrance requirement. Assuming that the scores are normally distributed, probabilities of achieving scores over various ranges of the GMAT can be determined. In a recent year, the mean GMAT score was 494 and the standard deviation was about 100. What is the probability that a randomly selected score from this administration of the GMAT is between 600 and the mean? That is,

$$P(494 \leq x \leq 600 \mid \mu = 494 \text{ and } \sigma = 100) = ?$$

Figure 6.6 is a graphical representation of this problem.

The z formula yields the number of standard deviations that the x value, 600, is away from the mean.

$$z = \frac{x - \mu}{\sigma} = \frac{600 - 494}{100} = \frac{106}{100} = 1.06$$

The z value of 1.06 reveals that the GMAT score of 600 is 1.06 standard deviations more than the mean. The z distribution values in Table 6.2 give the probability of a value being between this value of x and the mean. The whole-number and tenths-place portion of the z score appear in the first column of Table 6.2 (the 1.0 portion of this z score). Across the top of the table are the values of the hundredths-place portion of the z score. For this z score, the hundredths-place value is 6. The probability value in Table 6.2 for $z = 1.06$ is .3554. The shaded portion of the curve at the top of the table indicates that the probability value given *always* is the probability or area between an x value and the mean. In this particular example, that is the desired area. Thus the answer is that .3554 of the scores on the GMAT are between a score of 600 and the mean of 494. Figure 6.7(a) depicts graphically the solution in terms of x values. Figure 6.7(b) shows the solution in terms of z values.

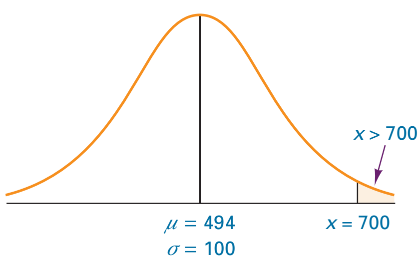
DEMONSTRATION PROBLEM 6.3

What is the probability of obtaining a score greater than 700 on a GMAT test that has a mean of 494 and a standard deviation of 100? Assume GMAT scores are normally distributed.

$$P(x > 700 \mid \mu = 494 \text{ and } \sigma = 100) = ?$$

Solution

Examine the following diagram.



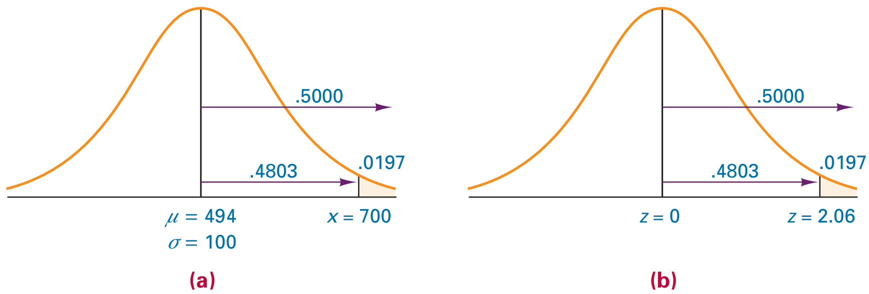
This problem calls for determining the area of the upper tail of the distribution. The z score for this problem is

$$z = \frac{x - \mu}{\sigma} = \frac{700 - 494}{100} = \frac{206}{100} = 2.06$$

Table 6.2 gives a probability of .4803 for this z score. This value is the probability of randomly drawing a GMAT with a score between the mean and 700. Finding the probability of getting a score greater than 700, which is the tail of the distribution, requires subtracting the probability value of .4803 from .5000, because each half of the distribution contains .5000 of the area. The result is .0197. Note that an attempt to determine the area of $x \geq 700$ instead of $x > 700$ would have made no difference because, in continuous distributions, the area under an exact number such as $x = 700$ is zero. A line segment has no width and hence no area.

.5000	(probability of x greater than the mean)
−.4803	(probability of x between 700 and the mean)
.0197	(probability of x greater than 700)

The solution is depicted graphically in (a) for x values and in (b) for z values.



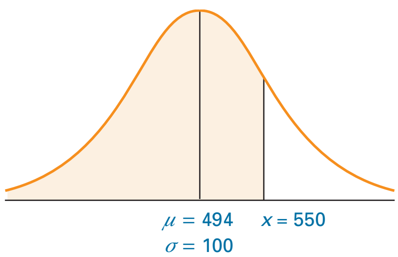
DEMONSTRATION PROBLEM 6.4

For the same GMAT examination, what is the probability of randomly drawing a score that is 550 or less?

$$P(x \leq 550 \mid \mu = 494 \text{ and } \sigma = 100) = ?$$

Solution

A sketch of this problem is shown here. Determine the area under the curve for all values less than or equal to 550.



The z value for the area between 550 and the mean is equal to:

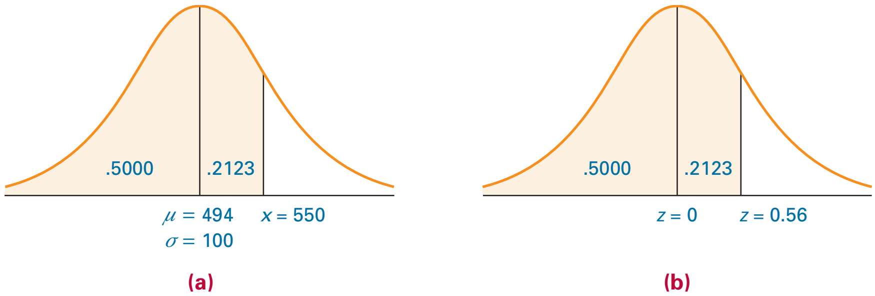
$$z = \frac{x - \mu}{\sigma} = \frac{550 - 494}{100} = \frac{56}{100} = 0.56$$

The area under the curve for $z = 0.56$ is .2123, which is the probability of getting a score between 550 and the mean. However, obtaining the probability for all values less than or equal to 550 also requires including the values less than the mean.

Because one-half or .5000 of the values are less than the mean, the probability of $x \leq 550$ is found as follows.

$$\begin{array}{rl} .5000 & \text{(probability of values less than the mean)} \\ \underline{+.2123} & \text{(probability of values between 550 and the mean)} \\ .7123 & \text{(probability of values } \leq 550\text{)} \end{array}$$

This solution is depicted graphically in (a) for x values and in (b) for z values.



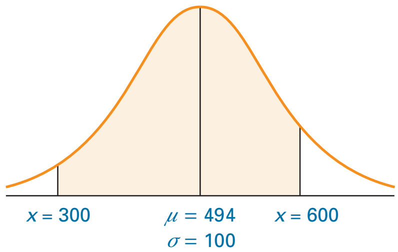
DEMONSTRATION PROBLEM 6.5

What is the probability of randomly obtaining a score between 300 and 600 on the GMAT exam?

$$P(300 < x < 600 \mid \mu = 494 \text{ and } \sigma = 100) = ?$$

Solution

The following sketch depicts the problem graphically: determine the area between $x = 300$ and $x = 600$, which spans the mean value. Because areas in the z distribution are given in relation to the mean, this problem must be worked as two separate problems and the results combined.



A z score is determined for each x value.

$$z = \frac{x - \mu}{\sigma} = \frac{600 - 494}{100} = \frac{106}{100} = 1.06$$

and

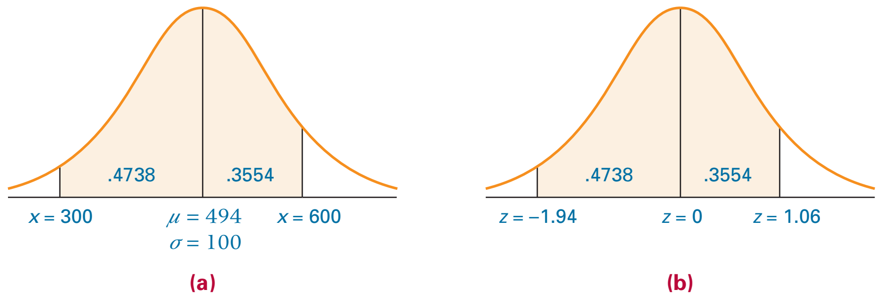
$$z = \frac{x - \mu}{\sigma} = \frac{300 - 494}{100} = \frac{-194}{100} = -1.94$$

Note that this z value ($z = -1.94$) is negative. A negative z value indicates that the x value is below the mean and the z value is on the left side of the distribution. None of the z values in Table 6.2 is negative. However, because the normal distribution is symmetric, probabilities for z values on the left side of the distribution are the same as the values on the right side of the distribution. The negative sign in the z value merely indicates that the area is on the left side of the distribution. The probability is always positive.

The probability for $z = 1.06$ is .3554; the probability for $z = -1.94$ is .4738. The solution of $P(300 < x < 600)$ is obtained by summing the probabilities.

.3554	(probability of a value between the mean and 600)
+.4738	(probability of a value between the mean and 300)
.8292	(probability of a value between 300 and 600)

Graphically, the solution is shown in (a) for x values and in (b) for z values.



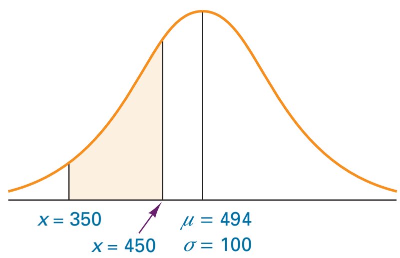
DEMONSTRATION PROBLEM 6.6

What is the probability of getting a score between 350 and 450 on the same GMAT exam?

$$P(350 < x < 450 \mid \mu = 494 \text{ and } \sigma = 100) = ?$$

Solution

The following sketch reveals that the solution to the problem involves determining the area of the shaded slice in the lower half of the curve.



In this problem, the two x values are on the same side of the mean. The areas or probabilities of each x value must be determined and the final probability found by determining the difference between the two areas.

$$z = \frac{x - \mu}{\sigma} = \frac{350 - 494}{100} = \frac{-144}{100} = -1.44$$

and

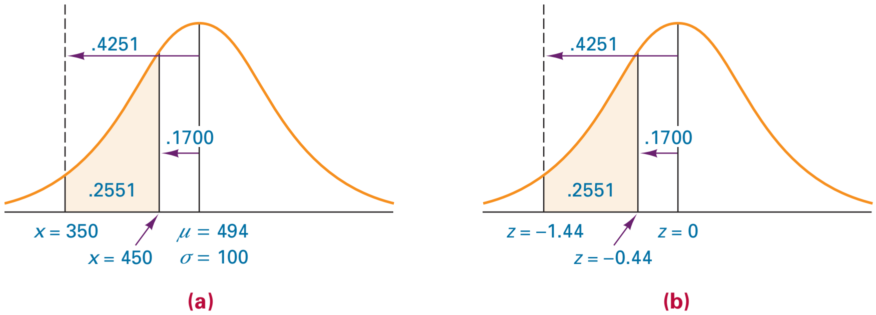
$$z = \frac{x - \mu}{\sigma} = \frac{450 - 494}{100} = \frac{-44}{100} = -0.44$$

The probability associated with $z = -1.44$ is .4251.
The probability associated with $z = -0.44$ is .1700.

Subtracting gives the solution.

.4251	(probability of a value between 350 and the mean)
−.1700	(probability of a value between 450 and the mean)
.2551	(probability of a value between 350 and 450)

Graphically, the solution is shown in (a) for x values and in (b) for z values.

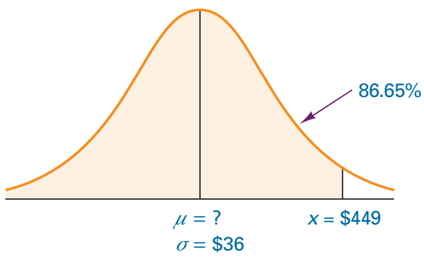


DEMONSTRATION PROBLEM 6.7

Runzheimer International publishes business travel costs for various cities throughout the world. In particular, they publish per diem totals, which represent the average costs for the typical business traveler including three meals a day in business-class restaurants and single-rate lodging in business-class hotels and motels. If 86.65% of the per diem costs in Buenos Aires, Argentina, are less than \$449 and if the standard deviation of per diem costs is \$36, what is the average per diem cost in Buenos Aires? Assume that per diem costs are normally distributed.

Solution

In this problem, the standard deviation and an x value are given; the object is to determine the value of the mean. Examination of the z score formula reveals four variables: x, μ, σ, and z. In this problem, only two of the four variables are given. Because solving one equation with two unknowns is impossible, one of the other unknowns must be determined. The value of z can be determined from the normal distribution table (Table 6.2).



Because 86.65% of the values are less than $x = \$449$, 36.65% of the per diem costs are between \$449 and the mean. The other 50% of the per diem costs are in the lower half of the distribution. Converting the percentage to a proportion yields .3665 of the values between the x value and the mean. What z value is associated with this area? This area, or probability, of .3665 in Table 6.2 is associated with the z value of 1.11. This z value is positive, because it is in the upper half of the distribution. Using the z value of 1.11, the x value of \$449, and the σ value of \$36 allows solving for the mean algebraically.

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

$$1.11 = \frac{\$449 - \mu}{\$36}$$

and

$$\mu = \$449 - (\$36)(1.11) = \$449 - \$39.96 = \$409.04$$

The mean per diem cost for business travel in Buenos Aires is \$409.04.

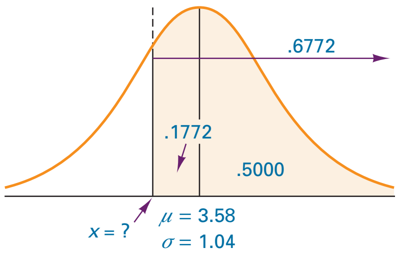
DEMONSTRATION PROBLEM 6.8

The U.S. Environmental Protection Agency publishes figures on solid waste generation in the United States. One year, the average number of waste generated per person per day was 3.58 pounds. Suppose the daily amount of waste generated per person is normally distributed, with a standard deviation of 1.04 pounds. Of the daily amounts of waste generated per person, 67.72% would be greater than what amount?

Solution

The mean and standard deviation are given, but x and z are unknown. The problem is to solve for a specific x value when .6772 of the x values are greater than that value.

If .6772 of the values are greater than x, then .1772 are between x and the mean (.6772 − .5000). Table 6.2 shows that the probability of.1772 is associated with a z value of 0.46. Because x is less than the mean, the z value actually is −0.46. Whenever an x value is less than the mean, its associated z value is negative and should be reported that way.



Solving the z equation yields

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

$$-0.46 = \frac{x - 3.58}{1.04}$$

and

$$x = 3.58 + (-0.46)(1.04) = 3.10$$

Thus 67.72% of the daily average amount of solid waste per person weighs more than 3.10 pounds.

STATISTICS IN BUSINESS TODAY

Warehousing

Tompkins Associates conducted a national study of warehousing in the United States. The study revealed many interesting facts. Warehousing is a labor-intensive industry that presents considerable opportunity for improvement in productivity. What does the "average" warehouse look like? The construction of new warehouses is restricted by prohibitive expense. Perhaps for that reason, the average age of a warehouse is 19 years. Warehouses vary in size, but the average size is about 50,000 square feet. To visualize such an "average" warehouse, picture one that is square with about 224 feet on each side or a rectangle that is 500 feet by 100 feet. The average clear height of a warehouse in the United States is 22 feet.

Suppose the ages of warehouses, the sizes of warehouses, and the clear heights of warehouses are normally distributed. Using the mean values already given and the standard deviations, techniques presented in this section could be used to determine, for example, the probability that a randomly selected warehouse is less than 15 years old, is larger than 60,000 square feet, or has a clear height between 20 and 25 feet.

TABLE 6.3
Excel and Minitab Normal Distribution Output for Demonstration Problem 6.6

Excel Output

x Value	
450	0.3264
350	0.0735
	0.2528

Minitab Output

CUMULATIVE DISTRIBUTION FUNCTION

Normal with mean = 494 and standard deviation = 100

x	$P(X <= x)$
450	0.329969
350	0.074934

Prob (350 < x < 450) = 0.255035

Using the Computer to Solve for Normal Distribution Probabilities

Both Excel and Minitab can be used to solve for normal distribution probabilities. In each case, the computer package uses μ, σ, and the value of x to compute a cumulative probability from the left. Shown in Table 6.3 are Excel and Minitab output for the probability question addressed in Demonstration Problem 6.6: $P(350 < x < 450 \mid \mu = 494 \text{ and } \sigma = 100)$. Since both computer packages yield probabilities cumulated from the left, this problem is solved manually with the computer output by finding the difference in $P(x < 450)$ and $P(x < 350)$.

6.2 PROBLEMS

6.6 Determine the probabilities for the following normal distribution problems.

a. $\mu = 604, \sigma = 56.8, x \leq 635$
b. $\mu = 48, \sigma = 12, x < 20$
c. $\mu = 111, \sigma = 33.8, 100 \leq x < 150$
d. $\mu = 264, \sigma = 10.9, 250 < x < 255$
e. $\mu = 37, \sigma = 4.35, x > 35$
f. $\mu = 156, \sigma = 11.4, x \geq 170$

6.7 Tompkins Associates reports that the mean clear height for a Class A warehouse in the United States is 22 feet. Suppose clear heights are normally distributed and that the standard deviation is 4 feet. A Class A warehouse in the United States is randomly selected.

a. What is the probability that the clear height is greater than 17 feet?
b. What is the probability that the clear height is less than 13 feet?
c. What is the probability that the clear height is between 25 and 31 feet?

6.8 According to a report by Scarborough Research, the average monthly household cellular phone bill is \$60. Suppose local monthly household cell phone bills are normally distributed with a standard deviation of \$11.35.

a. What is the probability that a randomly selected monthly cell phone bill is more than \$85?
b. What is the probability that a randomly selected monthly cell phone bill is between \$45 and \$70?
c. What is the probability that a randomly selected monthly cell phone bill is between \$65 and \$75?
d. What is the probability that a randomly selected monthly cell phone bill is no more than \$40?

6.9 According to the Internal Revenue Service, income tax returns one year averaged \$1,332 in refunds for taxpayers. One explanation of this figure is that taxpayers would rather have the government keep back too much money during the year than to owe it money at the end of the year. Suppose the average amount of tax at the end of a year is a refund of \$1,332, with a standard deviation of \$725. Assume that amounts owed or due on tax returns are normally distributed.

a. What proportion of tax returns show a refund greater than \$2,000?

b. What proportion of the tax returns show that the taxpayer owes money to the government?

c. What proportion of the tax returns show a refund between \$100 and \$700?

6.10 Toolworkers are subject to work-related injuries. One disorder, caused by strains to the hands and wrists, is called carpal tunnel syndrome. It strikes as many as 23,000 workers per year. The U.S. Labor Department estimates that the average cost of this disorder to employers and insurers is approximately \$30,000 per injured worker. Suppose these costs are normally distributed, with a standard deviation of \$9,000.

a. What proportion of the costs are between \$15,000 and \$45,000?

b. What proportion of the costs are greater than \$50,000?

c. What proportion of the costs are between \$5,000 and \$20,000?

d. Suppose the standard deviation is unknown, but 90.82% of the costs are more than \$7,000. What would be the value of the standard deviation?

e. Suppose the mean value is unknown, but the standard deviation is still \$9,000. How much would the average cost be if 79.95% of the costs were less than \$33,000?

6.11 Suppose you are working with a data set that is normally distributed, with a mean of 200 and a standard deviation of 47. Determine the value of x from the following information.

a. 60% of the values are greater than x.

b. x is less than 17% of the values.

c. 22% of the values are less than x.

d. x is greater than 55% of the values.

6.12 Suppose the annual employer 401(k) cost per participant is normally distributed with a standard deviation of \$625, but the mean is unknown.

a. If 73.89% of such costs are greater than \$1,700, what is the mean annual employer 401(k) cost per participant?

b. Suppose the mean annual employer 401(k) cost per participant is \$2,258 and the standard deviation is \$625. If such costs are normally distributed, 31.56% of the costs are greater than what value?

6.13 Suppose the standard deviation for Problem 6.7 is unknown but the mean is still 22 feet. If 72.4% of all U.S. Class A warehouses have a clear height greater than 18.5 feet, what is the standard deviation?

6.14 Suppose the mean clear height of all U.S. Class A warehouses is unknown but the standard deviation is known to be 4 feet. What is the value of the mean clear height if 29% of U.S. Class A warehouses have a clear height less than 20 feet?

6.15 Data accumulated by the National Climatic Data Center shows that the average wind speed in miles per hour for St. Louis, Missouri, is 9.7. Suppose wind speed measurements are normally distributed for a given geographic location. If 22.45% of the time the wind speed measurements are more than 11.6 miles per hour, what is the standard deviation of wind speed in St. Louis?

6.16 According to Student Monitor, a New Jersey research firm, the average cumulated college student loan debt for a graduating senior is \$25,760. Assume that the standard deviation of such student loan debt is \$5,684. Thirty percent of these graduating seniors owe more than what amount?

6.3 USING THE NORMAL CURVE TO APPROXIMATE BINOMIAL DISTRIBUTION PROBLEMS

Interactive Applet

For certain types of binomial distribution problems, the normal distribution can be used to approximate the probabilities. As sample sizes become large, binomial distributions approach the normal distribution in shape regardless of the value of p. This phenomenon occurs faster (for smaller values of n) when p is near .50. Figures 6.8 through 6.10 show three binomial distributions. Note in Figure 6.8 that even though the sample size, n, is only 10, the binomial graph bears a strong resemblance to a normal curve.

The graph in Figure 6.9 ($n = 10$ and $p = .20$) is skewed to the right because of the low p value and the small size. For this distribution, the expected value is only 2 and the probabilities pile up at $x = 0$ and 1. However, when n becomes large enough, as in the binomial distribution ($n = 100$ and $p = .20$) presented in Figure 6.10, the graph is relatively symmetric around the mean ($\mu = n \cdot p = 20$) because enough possible outcome values to the left of $x = 20$ allow the curve to fall back to the x-axis.

For large n values, the binomial distribution is cumbersome to analyze without a computer. Table A.2 goes only to $n = 25$. The normal distribution is a good approximation for binomial distribution problems for large values of n.

To work a binomial problem by the normal curve requires a translation process. The first part of this process is to convert the two parameters of a binomial distribution, n and p, to the two parameters of the normal distribution, μ and σ. This process utilizes formulas from Chapter 5:

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

After completion of this, a test must be made to determine whether the normal distribution is a good enough approximation of the binomial distribution:

Does the interval $\mu \pm 3\sigma$ lie between 0 and n?

Recall that the empirical rule states that approximately 99.7%, or almost all, of the values of a normal curve are within three standard deviations of the mean. For a normal curve approximation of a binomial distribution problem to be acceptable, all possible x values should be between 0 and n, which are the lower and upper limits, respectively, of a binomial distribution. If $\mu \pm 3\sigma$ is not between 0 and n, do *not* use the normal distribution to work a binomial problem because the approximation is not good enough. Upon demonstration that the normal curve is a good approximation for a binomial problem, the procedure continues. Another rule of thumb for determining when to use the normal curve to approximate a binomial problem is that the approximation is good enough if both $n \cdot p > 5$ and $n \cdot q > 5$.

The process can be illustrated in the solution of the binomial distribution problem.

$$P(x \geq 25 \mid n = 60 \text{ and } p = .30) = ?$$

Note that this binomial problem contains a relatively large sample size and that none of the binomial tables in Appendix A.2 can be used to solve the problem. This problem is a good candidate for use of the normal distribution.

Translating from a binomial problem to a normal curve problem gives

$$\mu = n \cdot p = (60)(.30) = 18 \text{ and } \sigma = \sqrt{n \cdot p \cdot q} = 3.55$$

The binomial problem becomes a normal curve problem.

$$P(x \geq 25 \mid \mu = 18 \text{ and } \sigma = 3.55) = ?$$

Next, the test is made to determine whether the normal curve sufficiently fits this binomial distribution to justify the use of the normal curve.

$$\mu \pm 3\sigma = 18 \pm 3(3.55) = 18 \pm 10.65$$
$$7.35 \leq \mu \pm 3\sigma \leq 28.65$$

This interval is between 0 and 60, so the approximation is sufficient to allow use of the normal curve. Figure 6.11 is a Minitab graph of this binomial distribution. Notice how

FIGURE 6.8

The Binomial Distribution for $n = 10$ and $p = .50$

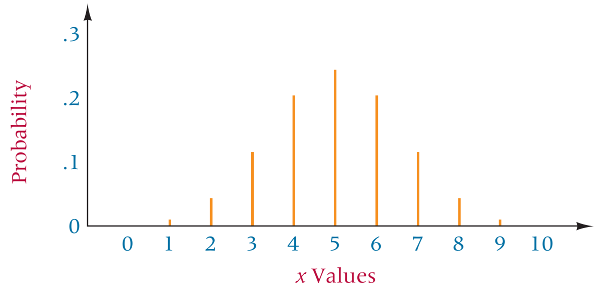


FIGURE 6.9

The Binomial Distribution for $n = 10$ and $p = .20$

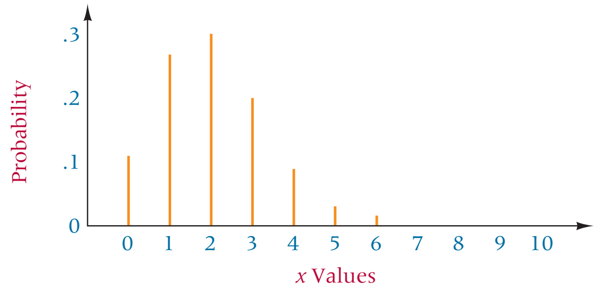


FIGURE 6.10

The Binomial Distribution for $n = 100$ and $p = .20$

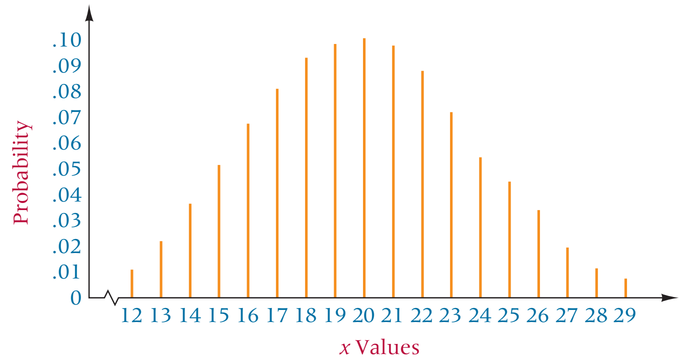
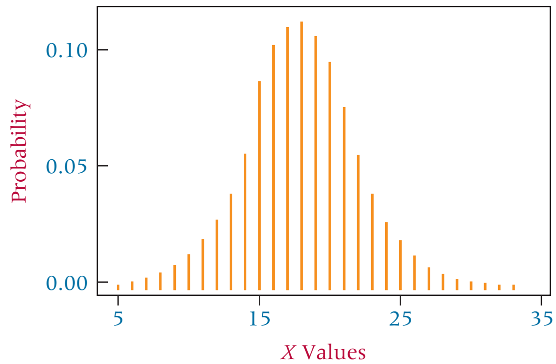


FIGURE 6.11

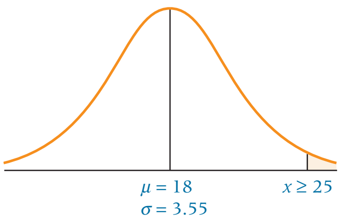
Graph of the Binomial Problem: $n = 60$ and $p = .30$



closely it resembles the normal curve. Figure 6.12 is the apparent graph of the normal curve version of this problem.

FIGURE 6.12

Graph of Apparent Solution of Binomial Problem Worked by the Normal Curve



Correcting for Continuity

The translation of a discrete distribution to a continuous distribution is not completely straightforward. A correction of +.50 or −.50 or ±.50, depending on the problem, is required. This correction ensures that most of the binomial problem's information is correctly transferred to the normal curve analysis. This correction is called the **correction for continuity**, which is *made during conversion of a discrete distribution into a continuous distribution.*

Figure 6.13 is a portion of the graph of the binomial distribution, $n = 60$ and $p = .30$. Note that with a binomial distribution, all the probabilities are concentrated on the whole numbers. Thus, the answers for $x \geq 25$ are found by summing the probabilities for $x = 25$, 26, 27, . . . , 60. There are no values between 24 and 25, 25 and 26, . . . , 59, and 60. Yet, the normal distribution is continuous, and values are present all along the x-axis. A correction must be made for this discrepancy for the approximation to be as accurate as possible.

As an analogy, visualize the process of melting iron rods in a furnace. The iron rods are like the probability values on each whole number of a binomial distribution. Note that the binomial graph in Figure 6.13 looks like a series of iron rods in a line. When the rods are placed in a furnace, they melt down and spread out. Each rod melts and moves to fill the area between it and the adjacent rods. The result is a continuous sheet of solid iron (continuous iron) that looks like the normal curve. The melting of the rods is analogous to spreading the binomial distribution to approximate the normal distribution.

How far does each rod spread toward the others? A good estimate is that each rod goes about halfway toward the adjacent rods. In other words, a rod that was concentrated at $x = 25$ spreads to cover the area from 24.5 to 25.5; $x = 26$ becomes continuous from 25.5 to 26.5; and so on. For the problem $P(x \geq 25 \mid n = 60 \text{ and } p = .30)$, conversion to a continuous normal curve problem yields $P(x \geq 24.5 \mid \mu = 18 \text{ and } \sigma = 3.55)$. The correction for continuity was −.50 because the problem called for the inclusion of the value of 25 along with all greater values; the binomial value of $x = 25$ translates to the normal curve value of 24.5 to 25.5. Had the binomial problem been to analyze $P(x > 25)$, the correction would have been +.50, resulting in a normal curve problem of $P(x \geq 25.5)$. The latter case would begin at more than 25 because the value of 25 would not be included.

TABLE 6.4

Rules of Thumb for the Correction for Continuity

Values Being Determined	Corrections
$x >$	+.50
$x \geq$	−.50
$x <$	−.50
$x \leq$	+.50
$\leq x \leq$	−.50 and +.50
$< x <$	+.50 and −.50
$x =$	−.50 and +.50

The decision as to how to correct for continuity depends on the equality sign and the direction of the desired outcomes of the binomial distribution. Table 6.4 lists some rules of thumb that can help in the application of the correction for continuity.

For the binomial problem $P(x \geq 25 \mid n = 60 \text{ and } p = .30)$, the normal curve becomes $P(x \geq 24.5 \mid \mu = 18 \text{ and } \sigma = 3.55)$, as shown in Figure 6.14, and

$$z = \frac{x - \mu}{\sigma} = \frac{24.5 - 18}{3.55} = 1.83$$

FIGURE 6.13

Graph of a Portion of the Binomial Problem: $n = 60$ and $p = .30$

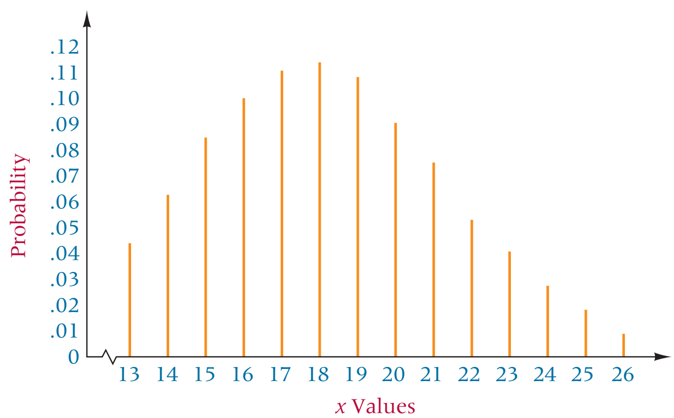


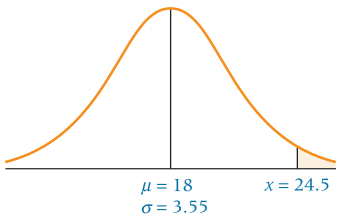
TABLE 6.5

Probability Values for the Binomial Problem: $n = 60$, $p = .30$, and $x \geq 25$

x Value	Probability
25	.0167
26	.0096
27	.0052
28	.0026
29	.0012
30	.0005
31	.0002
32	.0001
33	.0000
$x \geq 25$	.0361

FIGURE 6.14

Graph of the Solution to the Binomial Problem Worked by the Normal Curve



The probability (Table 6.2) of this z value is .4664. The answer to this problem lies in the tail of the distribution, so the final answer is obtained by subtracting.

$$\begin{array}{r} .5000 \\ \underline{-.4664} \\ .0336 \end{array}$$

Had this problem been worked by using the binomial formula, the solution would have been as shown in Table 6.5. The difference between the normal distribution approximation and the actual binomial values is only .0025 (.0361 − .0336).

DEMONSTRATION PROBLEM 6.9

Work the following binomial distribution problem by using the normal distribution.

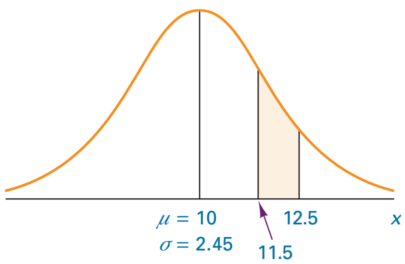
$$P(x = 12 \mid n = 25 \text{ and } p = .40) = ?$$

Solution

Find μ and σ.

$$\mu = n \cdot p = (25)(.40) = 10.0$$
$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{(25)(.40)(.60)} = 2.45$$
$$\text{test}: \mu \pm 3\sigma = 10.0 \pm 3(2.45) = 2.65 \text{ to } 17.35$$

This range is between 0 and 25, so the approximation is close enough. Correct for continuity next. Because the problem is to determine the probability of x being exactly 12, the correction entails both −.50 and + .50. That is, a binomial probability at $x = 12$ translates to a continuous normal curve area that lies between 11.5 and 12.5. The graph of the problem follows:



Then,

$$z = \frac{x - \mu}{\sigma} = \frac{12.5 - 10}{2.45} = 1.02$$

and

$$z = \frac{x - \mu}{\sigma} = \frac{11.5 - 10}{2.45} = 0.61$$

$z = 1.02$ produces a probability of .3461.

$z = 0.61$ produces a probability of .2291.

The difference in areas yields the following answer:

$$.3461 - .2291 = .1170$$

Had this problem been worked by using the binomial tables, the resulting answer would have been .114. The difference between the normal curve approximation and the value obtained by using binomial tables is only .003.

DEMONSTRATION PROBLEM 6.10

Solve the following binomial distribution problem by using the normal distribution.

$$P(x < 27 \mid n = 100 \text{ and } p = .37) = ?$$

Solution

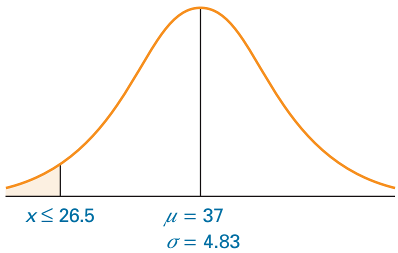
Because neither the sample size nor the p value is contained in Table A.2, working this problem by using binomial distribution techniques is impractical. It is a good candidate for the normal curve. Calculating μ and σ yields

$$\mu = n \cdot p = (100)(.37) = 37.0$$
$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{(100)(.37)(.63)} = 4.83$$

Testing to determine the closeness of the approximation gives

$$\mu \pm 3\sigma = 37 \pm 3(4.83) = 37 \pm 14.49$$

The range 22.51 to 51.49 is between 0 and 100. This problem satisfies the conditions of the test. Next, correct for continuity: $x < 27$ as a binomial problem translates to $x \le 26.5$ as a normal distribution problem. The graph of the problem follows.



Then,

$$z = \frac{x - \mu}{\sigma} = \frac{26.5 - 37}{4.83} = -2.17$$

Table 6.2 shows a probability of .4850. Solving for the tail of the distribution gives

$$.5000 - .4850 = .0150$$

which is the answer.

Had this problem been solved by using the binomial formula, the probabilities would have been the following.

x Value	Probability
26	.0059
25	.0035
24	.0019
23	.0010
22	.0005
21	.0002
20	.0001
$x < 27$	.0131

The answer obtained by using the normal curve approximation (.0150) compares favorably to this exact binomial answer. The difference is only .0019.

STATISTICS IN BUSINESS TODAY

Teleworking Facts

There are many interesting statistics about teleworkers. In a recent year, there were 45 million teleworkers in the United States, and more than 18% of employed adult Americans telework from home during business hours at least one day per month. Fifty-seven percent of HR professionals indicate that their organizations offer some form of telecommuting. The typical teleworker works an average of 5.5 days at home per month. The average commuting distance of a teleworker when he/she is not teleworking is 18 miles. Teleworkers save an average of 53 minutes commuting each day, saving them the equivalent of one extra day of work for every nine days of commuting. Thirty-three percent of Canadians would prefer to telework over a 10% wage increase, and 43% would change jobs to an employer allowing telework. Sixty-five percent of home teleworkers are males versus 44% of non-teleworkers. Among 20 United States, government agencies, the average per-user cost of setting up telecommuting is \$1,920. Telecommuting saves 840 million gallons of fuel annually in the United States, and telecommuting saves the equivalent of 9 to 14 billion kilowatt-hours of electricity per year—the same amount of energy used by roughly 1 million United States households every year.

Source: Telecommuting and Remote Work Statistics site at: http://www.suitecommute.com/Statistics.htm; and Telework Facts at: http://www.telcoa.org/id33_m.htm

6.3 PROBLEMS

6.17 Convert the following binomial distribution problems to normal distribution problems. Use the correction for continuity.

a. $P(x \leq 16 \mid n = 30 \text{ and } p = .70)$
b. $P(10 < x \leq 20) \mid n = 25 \text{ and } p = .50)$
c. $P(x = 22 \mid n = 40 \text{ and } p = .60)$
d. $P(x > 14 \mid n = 16 \text{ and } p = .45)$

6.18 Use the test $\mu \pm 3\sigma$ to determine whether the following binomial distributions can be approximated by using the normal distribution.

a. $n = 8$ and $p = .50$
b. $n = 18$ and $p = .80$
c. $n = 12$ and $p = .30$
d. $n = 30$ and $p = .75$
e. $n = 14$ and $p = .50$

6.19 Where appropriate, work the following binomial distribution problems by using the normal curve. Also, use Table A.2 to find the answers by using the binomial distribution and compare the answers obtained by the two methods.

a. $P(x = 8 \mid n = 25 \text{ and } p = .40) = ?$
b. $P(x \geq 13 \mid n = 20 \text{ and } p = .60) = ?$
c. $P(x = 7 \mid n = 15 \text{ and } p = .50) = ?$
d. $P(x < 3 \mid n = 10 \text{ and } p = .70) = ?$

6.20 The Zimmerman Agency conducted a study for Residence Inn by Marriott of business travelers who take trips of five nights or more. According to this study, 37% of these travelers enjoy sightseeing more than any other activity that they do not get to do as much at home. Suppose 120 randomly selected business travelers who take trips of five nights or more are contacted. What is the probability that fewer than 40 enjoy sightseeing more than any other activity that they do not get to do as much at home?

6.21 One study on managers' satisfaction with management tools reveals that 59% of all managers use self-directed work teams as a management tool. Suppose 70 managers selected randomly in the United States are interviewed. What is the probability that fewer than 35 use self-directed work teams as a management tool?

6.22 According to the Yankee Group, 53% of all cable households rate cable companies as good or excellent in quality transmission. Sixty percent of all cable households rate cable companies as good or excellent in having professional personnel. Suppose 300 cable households are randomly contacted.

a. What is the probability that more than 175 cable households rate cable companies as good or excellent in quality transmission?

b. What is the probability that between 165 and 170 (inclusive) cable households rate cable companies as good or excellent in quality transmission?

c. What is the probability that between 155 and 170 (inclusive) cable households rate cable companies as good or excellent in having professional personnel?

d. What is the probability that fewer than 200 cable households rate cable companies as good or excellent in having professional personnel?

6.23 Market researcher Gartner Dataquest reports that Dell Computer controls 27% of the PC market in the United States. Suppose a business researcher randomly selects 130 recent purchasers of PC.

a. What is the probability that more than 39 PC purchasers bought a Dell computer?

b. What is the probability that between 28 and 38 PC purchasers (inclusive) bought a Dell computer?

c. What is the probability that fewer than 23 PC purchasers bought a Dell computer?

d. What is the probability that exactly 33 PC purchasers bought a Dell computer?

6.24 A study about strategies for competing in the global marketplace states that 52% of the respondents agreed that companies need to make direct investments in foreign countries. It also states that about 70% of those responding agree that it is attractive to have a joint venture to increase global competitiveness. Suppose CEOs of 95 manufacturing companies are randomly contacted about global strategies.

a. What is the probability that between 44 and 52 (inclusive) CEOs agree that companies should make direct investments in foreign countries?

b. What is the probability that more than 56 CEOs agree with that assertion?

c. What is the probability that fewer than 60 CEOs agree that it is attractive to have a joint venture to increase global competitiveness?

d. What is the probability that between 55 and 62 (inclusive) CEOs agree with that assertion?

6.4 EXPONENTIAL DISTRIBUTION

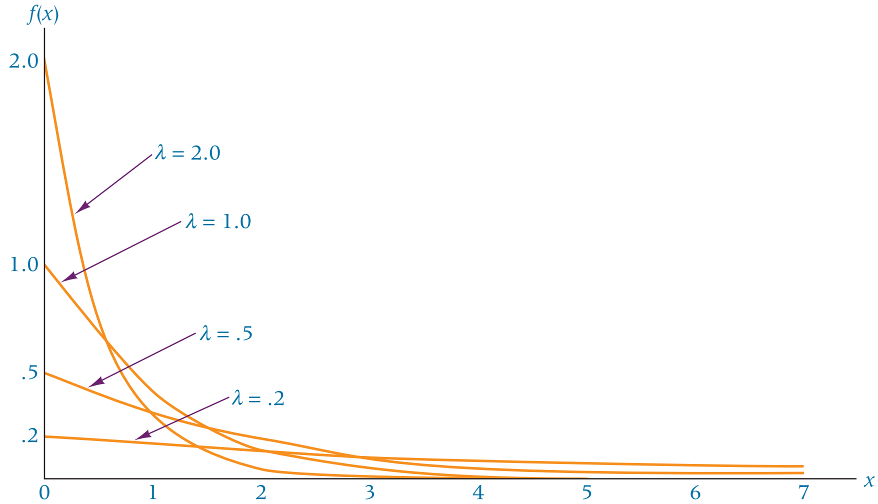
Another useful continuous distribution is the exponential distribution. It is closely related to the Poisson distribution. Whereas the Poisson distribution is discrete and describes random occurrences over some interval, the **exponential distribution** is *continuous and describes a probability distribution of the times between random occurrences.* The following are the characteristics of the exponential distribution.

- It is a continuous distribution.
- It is a family of distributions.
- It is skewed to the right.
- The x values range from zero to infinity.
- Its apex is always at $x = 0$.
- The curve steadily decreases as x gets larger.

The exponential probability distribution is determined by the following.

FIGURE 6.15

Graphs of Some Exponential Distributions



EXPONENTIAL PROBABILITY DENSITY FUNCTION

$$f(x) = \lambda e^{-\lambda x}$$

where

$x \geq 0$
$\lambda > 0$
and $e = 2.71828 \ldots$

An exponential distribution can be characterized by the one parameter, λ. Each unique value of λ determines a different exponential distribution, resulting in a family of exponential distributions. Figure 6.15 shows graphs of exponential distributions for four values of λ. The points on the graph are determined by using λ and various values of x in the probability density formula. The mean of an exponential distribution is $\mu = 1/\lambda$, and the standard deviation of an exponential distribution is $\sigma = 1/\lambda$.

Probabilities of the Exponential Distribution

Probabilities are computed for the exponential distribution by determining the area under the curve between two points. Applying calculus to the exponential probability density function produces a formula that can be used to calculate the probabilities of an exponential distribution.

PROBABILITIES OF THE RIGHT TAIL OF THE EXPONENTIAL DISTRIBUTION

$$P(x \geq x_0) = e^{-\lambda x_0}$$

where:

$x_0 \geq 0$

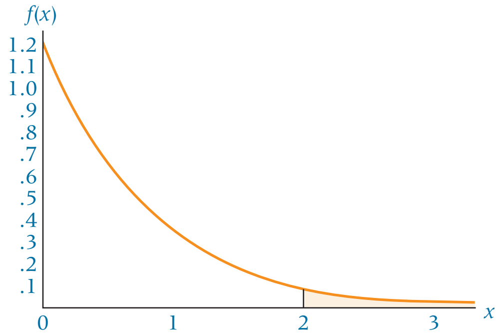
To use this formula requires finding values of e^{-x}. These values can be computed on most calculators or obtained from Table A.4, which contains the values of e^{-x} for selected values of x. x_0 is the fraction of the interval or the number of intervals between arrivals in the probability question and λ is the average arrival rate.

For example, arrivals at a bank are Poisson distributed with a λ of 1.2 customers every minute. What is the average time between arrivals and what is the probability that at least 2 minutes will elapse between one arrival and the next arrival? Since the interval for lambda is 1 minute and we want to know the probability that at least 2 minutes transpire between arrivals (twice the lambda interval), x_0 is 2.

Interarrival times of random arrivals are exponentially distributed. The mean of this exponential distribution is $\mu = 1/\lambda = 1/1.2 = .833$ minute (50 seconds). On average,

FIGURE 6.16

Exponential Distribution for $\lambda = 1.2$ and Solution for $x \geq 2$



.833 minute, or 50 seconds, will elapse between arrivals at the bank. The probability of an interval of 2 minutes or more between arrivals can be calculated by

$$P(x \geq 2 \mid \lambda = 1.2) = e^{-1.2(2)} = .0907.$$

About 9.07% of the time when the rate of random arrivals is 1.2 per minute, 2 minutes or more will elapse between arrivals, as shown in Figure 6.16.

This problem underscores the potential of using the exponential distribution in conjunction with the Poisson distribution to solve problems. In operations research and management science, these two distributions are used together to solve queuing problems (theory of waiting lines). The Poisson distribution can be used to analyze the arrivals to the queue, and the exponential distribution can be used to analyze the interarrival time.

DEMONSTRATION PROBLEM 6.11

A manufacturing firm has been involved in statistical quality control for several years. As part of the production process, parts are randomly selected and tested. From the records of these tests, it has been established that a defective part occurs in a pattern that is Poisson distributed on the average of 1.38 defects every 20 minutes during production runs. Use this information to determine the probability that less than 15 minutes will elapse between any two defects.

Solution

The value of λ is 1.38 defects per 20-minute interval. The value of μ can be determined by

$$\mu = \frac{1}{\lambda} = \frac{1}{1.38} = .7246$$

On the average, it is .7246 of the interval, or (.7246)(20 minutes) = 14.49 minutes, between defects. The value of x_0 represents the desired number of intervals between arrivals or occurrences for the probability question. In this problem, the probability question involves 15 minutes and the interval is 20 minutes. Thus x_0 is 15/20, or .75 of an interval. The question here is to determine the probability of there being less than 15 minutes between defects. The probability formula always yields the right tail of the distribution—in this case, the probability of there being 15 minutes or more between arrivals. By using the value of x_0 and the value of λ, the probability of there being 15 minutes or more between defects can be determined.

$$P(x \geq x_0) = P(x \geq .75) = e^{-\lambda x_0} = e^{(-1.38)(.75)} = e^{-1.035} = .3552$$

The probability of .3552 is the probability that at least 15 minutes will elapse between defects. To determine the probability of there being less than 15 minutes between defects, compute $1 - P(x)$. In this case, $1 - .3552 = .6448$. There is a probability of .6448 that less than 15 minutes will elapse between two defects when there is an average of 1.38 defects per 20-minute interval or an average of 14.49 minutes between defects.

TABLE 6.6

Excel and Minitab Output for Exponential Distribution

Excel Output

x Value	Probability < *x* Value
0.75	0.6448

Minitab Output

```
CUMULATIVE DISTRIBUTION FUNCTION

Exponential with mean = 0.7246
   x   P(X <= x)
0.75     0.644793
```

Using the Computer to Determine Exponential Distribution Probabilities

Both Excel and Minitab can be used to solve for exponential distribution probabilities. Excel uses the value of λ and x_0, but Minitab requires μ (equals $1/\lambda$) and x_0. In each case, the computer yields the cumulative probability from the left (the complement of what the probability formula shown in this section yields). Table 6.6 provides Excel and Minitab output for the probability question addressed in Demonstration Problem 6.11.

6.4 PROBLEMS

6.25 Use the probability density formula to sketch the graphs of the following exponential distributions.

a. $\lambda = 0.1$
b. $\lambda = 0.3$
c. $\lambda = 0.8$
d. $\lambda = 3.0$

6.26 Determine the mean and standard deviation of the following exponential distributions.

a. $\lambda = 3.25$
b. $\lambda = 0.7$
c. $\lambda = 1.1$
d. $\lambda = 6.0$

6.27 Determine the following exponential probabilities.

a. $P(x \geq 5 \mid \lambda = 1.35)$
b. $P(x < 3 \mid \lambda = 0.68)$
c. $P(x > 4 \mid \lambda = 1.7)$
d. $P(x < 6 \mid \lambda = 0.80)$

6.28 The average length of time between arrivals at a turnpike tollbooth is 23 seconds. Assume that the time between arrivals at the tollbooth is exponentially distributed.

a. What is the probability that a minute or more will elapse between arrivals?
b. If a car has just passed through the tollbooth, what is the probability that no car will show up for at least 3 minutes?

6.29 A busy restaurant determined that between 6:30 P.M. and 9:00 P.M. on Friday nights, the arrivals of customers are Poisson distributed with an average arrival rate of 2.44 per minute.

a. What is the probability that at least 10 minutes will elapse between arrivals?
b. What is the probability that at least 5 minutes will elapse between arrivals?
c. What is the probability that at least 1 minute will elapse between arrivals?
d. What is the expected amount of time between arrivals?

6.30 During the summer at a small private airport in western Nebraska, the unscheduled arrival of airplanes is Poisson distributed with an average arrival rate of 1.12 planes per hour.

a. What is the average interarrival time between planes?

b. What is the probability that at least 2 hours will elapse between plane arrivals?

c. What is the probability of two planes arriving less than 10 minutes apart?

6.31 The exponential distribution can be used to solve Poisson-type problems in which the intervals are not time. The Airline Quality Rating Study published by the U.S. Department of Transportation reported that in a recent year, Airtran led the nation in fewest occurrences of mishandled baggage, with a mean rate of 4.06 per 1,000 passengers. Assume mishandled baggage occurrences are Poisson distributed. Using the exponential distribution to analyze this problem, determine the average number of passengers between occurrences. Suppose baggage has just been mishandled.

a. What is the probability that at least 500 passengers will have their baggage handled properly before the next mishandling occurs?

b. What is the probability that the number will be fewer than 200 passengers?

6.32 The Foundation Corporation specializes in constructing the concrete foundations for new houses in the South. The company knows that because of soil types, moisture conditions, variable construction, and other factors, eventually most foundations will need major repair. On the basis of its records, the company's president believes that a new house foundation on average will not need major repair for 20 years. If she wants to guarantee the company's work against major repair but wants to have to honor no more than 10% of its guarantees, for how many years should the company guarantee its work? Assume that occurrences of major foundation repairs are Poisson distributed.

6.33 During the dry month of August, one U.S. city has measurable rain on average only two days per month. If the arrival of rainy days is Poisson distributed in this city during the month of August, what is the average number of days that will pass between measurable rain? What is the standard deviation? What is the probability during this month that there will be a period of less than two days between rain?

The Cost of Human Resources

Decision Dilemma Solved

The American Society for Training and Development reported that, on average, an employee receives 32 hours of training per year. Suppose that number of hours of training is uniformly distributed across all employees varying from 0 hours to 64 hours. Using techniques presented in Section 6.1, this uniform distribution can be described by $a = 0$, $b = 64$, and $\mu = 32$. The probability that an employee receives between 20 and 40 hours of training can be determined by the following calculation assuming that $x_1 = 20$ and $x_2 = 40$:

$$P(x) = \frac{x_2 - x_1}{b - a} = \frac{40 - 20}{64 - 0} = \frac{20}{64} = .3125$$

Thus, 31.25% of employees receive between 20 and 40 hours of training.

The probability that an employee receives 50 hours or more of training can be calculated as:

$$P(x) = \frac{x_2 - x_1}{b - a} = \frac{64 - 50}{64 - 0} = \frac{14}{64} = .21875$$

Almost 22% of employees receive 50 hours or more of training. Note that here, x_2 is 64 since 64 hours is the upper end of the distribution.

It is estimated by some studies that, on average, it costs \$3,270 to hire an employee. If such costs are normally distributed with a standard deviation of \$400, the probability that it costs more than \$4,000 to hire an employee can be calculated using techniques from Section 6.2 as:

$$z = \frac{x - \mu}{\sigma} = \frac{4000 - 3270}{400} = 1.83$$

The area associated with this z value is .4664 and the tail of the distribution is $.5000 - .4664 = .0336$. That is, 3.36% of the time, it costs more than $4,000 to hire an employee. The probability that it costs less than $3,000 to hire an employee can be determined in a similar manner:

$$z = \frac{x - \mu}{\sigma} = \frac{3000 - 3270}{400} = 0.68$$

The area associated with this z value is .2517 and the tail of the distribution is $.5000 - .2517 = .2483$. That is, 24.83% of the time, it costs less than $3,000 to hire an employee.

Thirty-five percent of all unscheduled absenteeism is caused by personal illness. Using techniques presented in Section 6.3, the probability that more than 50 of 120 randomly selected unscheduled absences were caused by personal illness can be determined. With $n = 120$, $p = .35$ and $x > 50$, this binomial distribution problem can be converted into a normal distribution problem by:

$$\mu = n \cdot p = (120)(.35) = 42$$

and

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{(120)(.35)(.65)} = 5.225$$

Since $42 \pm 3(5.225)$ is between 0 and 120, it is appropriate to use the normal distribution to approximate this binomial problem. Applying the correction for continuity, $x \geq 50.5$. The z value is calculated as:

$$z = \frac{x - \mu}{\sigma} = \frac{50.5 - 42}{5.225} = 1.63$$

The area associated with this z value is .4484 and the tail of the distribution is $.5000 - .4484 = .0516$. That is, 5.16% of the time, more than 50 out of 120 unscheduled absences are due to personal illness.

ETHICAL CONSIDERATIONS

Several points must be considered in working with continuous distributions. Is the population being studied the same population from which the parameters (mean, standard deviation, λ) were determined? If not, the results may not be valid for the analysis being done. Invalid or spurious results can be obtained by using the parameters from one population to analyze another population. For example, a market study in New England may result in the conclusion that the amount of fish eaten per month by adults is normally distributed with the average of 2.3 pounds of fish per month. A market researcher in the Southwest should not assume that these figures apply to her population. People in the Southwest probably have quite different fish-eating habits than people in New England, and the application of New England population parameters to the Southwest probably will result in questionable conclusions.

As was true with the Poisson distribution in Chapter 5, the use of λ in the exponential distribution should be judicious because a λ for one interval in a given time period or situation may not be the same as the λ for the same interval in a different time period or situation. For example, the number of arrivals per five-minute time period at a restaurant on Friday night is not likely to be the same as the number of arrivals in a five-minute time period at that same restaurant from 2 P.M. to 4 P.M. on weekdays. In using established parameters such as μ and λ, a researcher should be certain that the population from which the parameter was determined is, indeed, the same population being studied.

Sometimes a normal distribution is used to analyze data when, in fact, the data are not normal. Such an analysis can contain bias and produce false results. Certain techniques for testing a distribution of data can determine whether they are distributed a certain way. Some of the techniques are presented in Chapter 16. In general, Chapter 6 techniques can be misused if the wrong type of distribution is applied to the data or if the distribution used for analysis is the right one but the parameters (μ, σ, λ) do not fit the data of the population being analyzed. ›

SUMMARY

This chapter discussed three different continuous distributions: the uniform distribution, the normal distribution, and the exponential distribution. With continuous distributions, the value of the probability density function does not yield the probability but instead gives the height of the curve at any given point. In fact, with continuous distributions, the probability at any discrete point is .0000. Probabilities are determined over an interval. In each case, the probability is the area under the curve for the interval being considered. In each distribution, the probability or total area under the curve is 1.

Probably the simplest of these distributions is the uniform distribution, sometimes referred to as the rectangular distribution. The uniform distribution is determined from a probability density function that contains equal values along some interval between the points a and b. Basically, the height of the curve is the same everywhere between these two points. Probabilities are determined by calculating the portion of the rectangle between the two points a and b that is being considered.

The most widely used of all distributions is the normal distribution. Many phenomena are normally distributed, including characteristics of most machine-produced parts, many measurements of the biological and natural environment, and many human characteristics such as height, weight, IQ, and achievement test scores. The normal curve is continuous, symmetrical, unimodal, and asymptotic to the axis; actually, it is a family of curves.

The parameters necessary to describe a normal distribution are the mean and the standard deviation. For convenience, data that are being analyzed by the normal curve should be standardized by using the mean and the standard deviation to compute z scores. A z score is the distance that an x value is from the mean, μ, in units of standard deviations. With the z score of an x value, the probability of that value occurring by chance from a given normal distribution can be determined by using a table of z scores and their associated probabilities.

The normal distribution can be used to work certain types of binomial distribution problems. Doing so requires converting the n and p values of the binomial distribution to μ and σ of the normal distribution. When worked by using the normal distribution, the binomial distribution solution is only an approximation. If the values of $\mu \pm 3\sigma$ are within a range from 0 to n, the approximation is reasonably accurate. Adjusting for the fact that a discrete distribution problem is being worked by using a continuous distribution requires a correction for continuity. The correction for continuity involves adding or subtracting .50 to the x value being analyzed. This correction usually improves the normal curve approximation.

Another continuous distribution is the exponential distribution. It complements the discrete Poisson distribution. The exponential distribution is used to compute the probabilities of times between random occurrences. The exponential distribution is a family of distributions described by one parameter, λ. The distribution is skewed to the right and always has its highest value at $x = 0$.

KEY TERMS

Flash Cards

correction for continuity
exponential distribution
normal distribution
rectangular distribution
standardized normal distribution
uniform distribution
z distribution
z score

FORMULAS

Probability density function of a uniform distribution

$$f(x) = \begin{cases} \dfrac{1}{b - a} & \text{for } a \le x \le b \\ 0 & \text{for all other values} \end{cases}$$

Mean and standard deviation of a uniform distribution

$$\mu = \frac{a + b}{2}$$

$$\sigma = \frac{b - a}{\sqrt{12}}$$

Probability density function of the normal distribution

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

z formula

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

Conversion of a binomial problem to the normal curve

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

Exponential probability density function

$$f(x) = \lambda e^{-\lambda x}$$

Probabilities of the right tail of the exponential distribution

$$P(x \ge x_0) = e^{-\lambda x_0}$$

SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

6.34 Data are uniformly distributed between the values of 6 and 14. Determine the value of $f(x)$. What are the mean and standard deviation of this distribution? What is the probability of randomly selecting a value greater than 11? What is the probability of randomly selecting a value between 7 and 12?

6.35 Assume a normal distribution and find the following probabilities.

a. $P(x < 21 \mid \mu = 25 \text{ and } \sigma = 4)$
b. $P(x \ge 77 \mid \mu = 50 \text{ and } \sigma = 9)$
c. $P(x > 47 \mid \mu = 50 \text{ and } \sigma = 6)$
d. $P(13 < x < 29 \mid \mu = 23 \text{ and } \sigma = 4)$
e. $P(x \ge 105 \mid \mu = 90 \text{ and } \sigma = 2.86)$

6.36 Work the following binomial distribution problems by using the normal distribution. Check your answers by using Table A.2 to solve for the probabilities.

a. $P(x = 12 \mid n = 25 \text{ and } p = .60)$

b. $P(x > 5 \mid n = 15 \text{ and } p = .50)$

c. $P(x \leq 3 \mid n = 10 \text{ and } p = .50)$

d. $P(x \geq 8 \mid n = 15 \text{ and } p = .40)$

6.37 Find the probabilities for the following exponential distribution problems.

a. $P(x \geq 3 \mid \lambda = 1.3)$

b. $P(x < 2 \mid \lambda = 2.0)$

c. $P(1 \leq x \leq 3 \mid \lambda = 1.65)$

d. $P(x > 2 \mid \lambda = .405)$

TESTING YOUR UNDERSTANDING

6.38 The U.S. Bureau of Labor Statistics reports that of persons who usually work full-time, the average number of hours worked per week is 43.4. Assume that the number of hours worked per week for those who usually work full-time is normally distributed. Suppose 12% of these workers work more than 48 hours. Based on this percentage, what is the standard deviation of number of hours worked per week for these workers?

6.39 A U.S. Bureau of Labor Statistics survey showed that one in five people 16 years of age or older volunteers some of his or her time. If this figure holds for the entire population and if a random sample of 150 people 16 years of age or older is taken, what is the probability that more than 50 of those sampled do volunteer work?

6.40 An entrepreneur opened a small hardware store in a strip mall. During the first few weeks, business was slow, with the store averaging only one customer every 20 minutes in the morning. Assume that the random arrival of customers is Poisson distributed.

a. What is the probability that at least one hour would elapse between customers?

b. What is the probability that 10 to 30 minutes would elapse between customers?

c. What is the probability that less than five minutes would elapse between customers?

6.41 In a recent year, the average price of a Microsoft Windows Upgrade was \$90.28 according to *PC Data*. Assume that prices of the Microsoft Windows Upgrade that year were normally distributed, with a standard deviation of \$8.53. If a retailer of computer software was randomly selected that year:

a. What is the probability that the price of a Microsoft Windows Upgrade was below \$80?

b. What is the probability that the price was above \$95?

c. What is the probability that the price was between \$83 and \$87?

6.42 According to the U.S. Department of Agriculture, Alabama egg farmers produce millions of eggs every year. Suppose egg production per year in Alabama is normally distributed, with a standard deviation of 83 million eggs. If during only 3% of the years Alabama egg farmers produce more than 2,655 million eggs, what is the mean egg production by Alabama farmers?

6.43 The U.S. Bureau of Labor Statistics releases figures on the number of full-time wage and salary workers with flexible schedules. The numbers of full-time wage and salary workers in each age category are almost uniformly distributed by age, with ages ranging from 18 to 65 years. If a worker with a flexible schedule is randomly drawn from the U.S. workforce, what is the probability that he or she will be between 25 and 50 years of age? What is the mean value for this distribution? What is the height of the distribution?

6.44 A business convention holds its registration on Wednesday morning from 9:00 A.M. until 12:00 noon. Past history has shown that registrant arrivals follow a Poisson distribution at an average rate of 1.8 every 15 seconds. Fortunately, several facilities are available to register convention members.

a. What is the average number of seconds between arrivals to the registration area for this conference based on past results?

b. What is the probability that 25 seconds or more would pass between registration arrivals?

c. What is the probability that less than five seconds will elapse between arrivals?

d. Suppose the registration computers went down for a one-minute period. Would this condition pose a problem? What is the probability that at least one minute will elapse between arrivals?

6.45 *M/PF Research, Inc.* lists the average monthly apartment rent in some of the most expensive apartment rental locations in the United States. According to their report, the average cost of renting an apartment in Minneapolis is \$951. Suppose that the standard deviation of the cost of renting an apartment in Minneapolis is \$96 and that apartment rents in Minneapolis are normally distributed. If a Minneapolis apartment is randomly selected, what is the probability that the price is:

a. \$1,000 or more?

b. Between \$900 and \$1,100?

c. Between \$825 and \$925?

d. Less than \$700?

6.46 According to *The Wirthlin Report*, 24% of all workers say that their job is very stressful. If 60 workers are randomly selected:

a. What is the probability that 17 or more say that their job is very stressful?

b. What is the probability that more than 22 say that their job is very stressful?

c. What is the probability that between 8 and 12 (inclusive) say that their job is very stressful?

6.47 The U.S. Bureau of Economic Statistics reports that the average annual salary in the metropolitan Boston area is \$50,542. Suppose annual salaries in the metropolitan Boston area are normally distributed with a standard deviation of \$4,246. A Boston worker is randomly selected.

a. What is the probability that the worker's annual salary is more than \$60,000?

b. What is the probability that the worker's annual salary is less than \$45,000?

c. What is the probability that the worker's annual salary is more than \$40,000?

d. What is the probability that the worker's annual salary is between \$44,000 and \$52,000?

6.48 Suppose interarrival times at a hospital emergency room during a weekday are exponentially distributed, with an average interarrival time of nine minutes. If the arrivals are Poisson distributed, what would the average number of arrivals per hour be? What is the probability that less than five minutes will elapse between any two arrivals?

6.49 Suppose the average speeds of passenger trains traveling from Newark, New Jersey, to Philadelphia, Pennsylvania, are normally distributed, with a mean average speed of 88 miles per hour and a standard deviation of 6.4 miles per hour.

a. What is the probability that a train will average less than 70 miles per hour?

b. What is the probability that a train will average more than 80 miles per hour?

c. What is the probability that a train will average between 90 and 100 miles per hour?

6.50 The Conference Board published information on why companies expect to increase the number of part-time jobs and reduce full-time positions. Eighty-one percent of the companies said the reason was to get a flexible workforce. Suppose 200 companies that expect to increase the number of part-time jobs and reduce full-time positions are identified and contacted. What is the expected number of these companies that would agree that the reason is to get a flexible workforce? What is the probability that between 150 and 155 (not including the 150 or the 155) would give that reason? What is the probability that more than 158 would give that reason? What is the probability that fewer than 144 would give that reason?

6.51 According to the U.S. Bureau of the Census, about 75% of commuters in the United States drive to work alone. Suppose 150 U.S. commuters are randomly sampled.

Demonstration Problem

a. What is the probability that fewer than 105 commuters drive to work alone?

b. What is the probability that between 110 and 120 (inclusive) commuters drive to work alone?

c. What is the probability that more than 95 commuters drive to work alone?

6.52 According to figures released by the National Agricultural Statistics Service of the U.S. Department of Agriculture, the U.S. production of wheat over the past 20 years has been approximately uniformly distributed. Suppose the mean production over this period was 2.165 billion bushels. If the height of this distribution is .862 billion bushels, what are the values of a and b for this distribution?

6.53 The Federal Reserve System publishes data on family income based on its Survey of Consumer Finances. When the head of the household has a college degree, the mean before-tax family income is \$85,200. Suppose that 60% of the before-tax family incomes when the head of the household has a college degree are between \$75,600 and \$94,800 and that these incomes are normally distributed. What is the standard deviation of before-tax family incomes when the head of the household has a college degree?

6.54 According to the Polk Company, a survey of households using the Internet in buying or leasing cars reported that 81% were seeking information about prices. In addition, 44% were seeking information about products offered. Suppose 75 randomly selected households who are using the Internet in buying or leasing cars are contacted.

a. What is the expected number of households who are seeking price information?

b. What is the expected number of households who are seeking information about products offered?

c. What is the probability that 67 or more households are seeking information about prices?

d. What is the probability that fewer than 23 households are seeking information about products offered?

6.55 Coastal businesses along the Gulf of Mexico from Texas to Florida worry about the threat of hurricanes during the season from June through October. Businesses become especially nervous when hurricanes enter the Gulf of Mexico. Suppose the arrival of hurricanes during this season is Poisson distributed, with an average of three hurricanes entering the Gulf of Mexico during the five-month season. If a hurricane has just entered the Gulf of Mexico:

a. What is the probability that at least one month will pass before the next hurricane enters the Gulf?

b. What is the probability that another hurricane will enter the Gulf of Mexico in two weeks or less?

c. What is the average amount of time between hurricanes entering the Gulf of Mexico?

6.56 With the growing emphasis on technology and the changing business environment, many workers are discovering that training such as reeducation, skill development, and personal growth are of great assistance in the job marketplace. A recent Gallup survey found that 80% of Generation Xers considered the availability of company-sponsored training as a factor to weigh in taking a job. If 50 Generation Xers are randomly sampled, what is the probability that fewer than 35 consider the availability of company-sponsored training as a factor to weigh in taking a job? What is the expected number? What is the probability that between 42 and 47 (inclusive) consider the availability of company-sponsored training as a factor to weigh in taking a job?

6.57 According to the Air Transport Association of America, the average operating cost of an MD-80 jet airliner is $2,087 per hour. Suppose the operating costs of an MD-80 jet airliner are normally distributed with a standard deviation of $175 per hour. At what operating cost would only 20% of the operating costs be less? At what operating cost would 65% of the operating costs be more? What operating cost would be more than 85% of operating costs?

6.58 Supermarkets usually become busy at about 5 P.M. on weekdays, because many workers stop by on the way home to shop. Suppose at that time arrivals at a supermarket's express checkout station are Poisson distributed, with an average of .8 person/minute. If the clerk has just checked out the last person in line, what is the probability that at least one minute will elapse before the next customer arrives? Suppose the clerk wants to go to the manager's office to ask a quick question and needs 2.5 minutes to do so. What is the probability that the clerk will get back before the next customer arrives?

6.59 In a recent year, the average daily circulation of *The Wall Street Journal* was 1,717,000. Suppose the standard deviation is 50,940. Assume the paper's daily circulation is normally distributed. On what percentage of days would circulation pass 1,800,000? Suppose the paper cannot support the fixed expenses of a full-production setup if the circulation drops below 1,600,000. If the probability of this even occurring is low, the production manager might try to keep the full crew in place and not disrupt operations. How often will this even happen, based on this historical information?

6.60 Incoming phone calls generally are thought to be Poisson distributed. If an operator averages 2.2 phone calls every 30 seconds, what is the expected (average) amount of time between calls? What is the probability that a minute or more would elapse between incoming calls? Two minutes?

INTERPRETING THE OUTPUT

6.61 Shown here is a Minitab output. Suppose the data represent the number of sales associates who are working in a department store in any given retail day. Describe the distribution including the mean and standard deviation. Interpret the shape of the distribution and the mean in light of the data being studied. What do the probability statements mean?

```
CUMULATIVE DISTRIBUTION FUNCTION

Continuous uniform on 11 to 32
x            P(X <= x)
28            0.80952
34            1.00000
16            0.23810
21            0.47619
```

6.62 A manufacturing company produces a metal rod. Use the Excel output shown here to describe the weight of the rod. Interpret the probability values in terms of the manufacturing process.

Normal Distribution
Mean = 227 mg.
Standard Deviation = 2.3 mg.

x Value	Probability < *x* Value
220	0.0012
225	0.1923
227	0.5000
231	0.9590
238	1.0000

6.63 Suppose the Minitab output shown here represents the analysis of the length of home-use cell phone calls in terms of minutes. Describe the distribution of cell phone call lengths and interpret the meaning of the probability statements.

```
CUMULATIVE DISTRIBUTION FUNCTION

Normal with mean = 2.35 and
standard deviation = 0.11
  x            P(X <= x)
2.60           0.988479
2.45           0.818349
2.30           0.324718
2.00           0.000732
```

6.64 A restaurant averages 4.51 customers per 10 minutes during the summer in the late afternoon. Shown here are Excel and Minitab output for this restaurant. Discuss the type of distribution used to analyze the data and the meaning of the probabilities.

Exponential Distribution

x Value	Probability < *x* Value
0.1	0.3630
0.2	0.5942
0.5	0.8951
1.0	0.9890
2.4	1.0000

```
CUMULATIVE DISTRIBUTION FUNCTION

Exponential with mean = 0.221729
x              P(X <= x)
0.1            0.363010
0.2            0.594243
0.5            0.895127
1.0            0.989002
2.4            0.999980
```

ANALYZING THE DATABASES

see www.wiley.com/college/black

1. The Consumer Food database contains a variable, Annual Food Spending, which represents the amount spent per household on food for a year. Calculate the mean and standard deviation for this variable that is approximately normally distributed in this database. Using the mean and standard deviation, calculate the probability that a randomly selected household spends more than $10,000 annually on food. What is the probability that a randomly selected household spends less than $5,000 annually on food? What is the probability that a randomly selected household spends between $8,000 and $11,000 annually on food?
2. Select the Agribusiness time-series database. Create a histogram graph for onions and for broccoli. Each of these variables is approximately normally distributed. Compute the mean and the standard deviation for each distribution. The data in this database represent the monthly weight (in thousands of pounds) of each vegetable. In terms of monthly weight, describe each vegetable (onions and broccoli). If a month were randomly selected from the onion distribution, what is the probability that the weight would be more than 50,000? What is the probability that the weight would be between 25,000 and 35,000? If a month were randomly selected from the broccoli distribution, what is the probability that the weight would be more than 100,000? What is the probability that the weight would be between 135,000 and 170,000?
3. From the Hospital database, it can be determined that some hospitals admit around 50 patients per day. Suppose we select a hospital that admits 50 patients per day. Assuming that admittance only occurs within a 12-hour time period each day and that admittance is Poisson distributed, what is the value of lambda for per hour for this hospital? What is the interarrival time for admittance based on this figure? Suppose a person was just admitted to the hospital. What the probability that it would be more than 30 minutes before the next person was admitted? What is the probability that there would be less than 10 minutes before the next person was admitted?

CASE

MERCEDES GOES AFTER YOUNGER BUYERS

Mercedes and BMW have been competing head-to-head for market share in the luxury-car market for more than four decades. Back in 1959, BMW (Bayerische Motoren Werke) almost went bankrupt and nearly sold out to Daimler-Benz, the maker of Mercedes-Benz cars. BMW was able to recover to the point that in 1992 it passed Mercedes in worldwide sales. Among the reasons for BMW's success was its ability to sell models that were more luxurious than previous models but still focused on consumer quality and environmental responsibility. In particular, BMW targeted its sales pitch to the younger market, whereas Mercedes retained a more mature customer base.

In response to BMW's success, Mercedes has been trying to change their image by launching several products in an effort to attract younger buyers who are interested in sporty, performance-oriented cars. BMW, influenced by Mercedes, is pushing for more refinement and comfort. In fact, one automotive expert says that Mercedes wants to become BMW, and vice versa. However, according to one recent automotive expert, the focus is still on luxury and comfort for Mercedes while BMW focuses on performance and driving dynamics. Even though each company produces many different models, two relatively comparable coupe automobiles are the BMW 3 Series Coupe 335i and the Mercedes CLK350 Coupe. In a recent year, the national U.S. market price for the BMW 3 Series Coupe 335i was $39,368 and for the Mercedes CLK350 Couple was $44,520. Gas mileage for both of these cars is around 17 mpg in town and 25 mpg on the highway.

Discussion

1. Suppose Mercedes is concerned that dealer prices of the CLK350 Coupe are not consistent and that even though the average price is $44,520, actual prices are normally distributed with a standard deviation of $2,981. Suppose also that Mercedes believes that at $43,000, the CLK350 Coupe is priced out of the BMW 3 Series Coupe 335i market. What percentage of the dealer prices for the Mercedes CLK350 Coupe is more than $43,000 and hence priced out of the BMW 3 Series Coupe 335i market? The average price for a BMW 3 Series Coupe 335i is $39,368. Suppose these prices are also normally distributed with a standard deviation of $2,367. What percentage of BMW dealers are pricing the BMW 3 Series Coupe 335i at more than the average price for a Mercedes CLK350 Coupe? What might this mean to BMW if dealers were pricing the 3 Series Couple 351 at this level? What percentage of Mercedes dealers are pricing the CLK350 Couple at less than the average price of a BMW 3 Series Coupe 335i?
2. Suppose that highway gas mileage rates for both of these cares are uniformly distributed over a range of from 20 to 30 mpg. What proportion of these cars would fall into the 22 to 27 mpg range? Compute the proportion of cars that get more than 28 mpg. What proportion of cars would get less than 23 mpg?
3. Suppose that in one dealership an average of 1.37 CLKs is sold every 3 hours (during a 12-hour showroom day) and

that sales are Poisson distributed. The following Excel-produced probabilities indicate the occurrence of different intersales times based on this information. Study the output and interpret it for the salespeople. For example, what is the probability that less than an hour will elapse between sales? What is the probability that more than a day (12-hour day) will pass before the next sale after a car has been sold? What can the dealership managers do with such information? How can it help in staffing? How can such information be used as a tracking device for the impact of advertising? Is there a chance that these probabilities would change during the year? If so, why?

Portion of 3-Hour Time Frame	Cumulative Exponential Probabilities from Left
0.167	0.2045
0.333	0.3663
0.667	0.5990
1	0.7459
2	0.9354
3	0.9836
4	0.9958
5	0.9989

USING THE COMPUTER

EXCEL

- Excel can be used to compute cumulative probabilities for particular values of *x* from either a normal distribution or an exponential distribution.
- Calculation of probabilities from each of these distributions begins with the **Insert Function** (f_x). To access the **Insert Function**, go to the **Formulas** tab on an Excel worksheet (top center tab). The **Insert Function** is on the far left of the menu bar. In the **Insert Function** dialog box at the top, there is a pulldown menu where it says **Or select a category**. From the pulldown menu associated with this command, select **Statistical**.
- To compute probabilities from a normal distribution, select **NORMDIST** from the **Insert Function's Statistical** menu. In the **NORMDIST** dialog box, there are four lines to which you must respond. On the first line, **X**, enter the value of *x*. On the second line, **Mean**, enter the value of the mean. On the third line, **Standard_dev**, enter the value of the standard deviation. The fourth line, **Cumulative**, requires a logical response of either TRUE or FALSE. Place TRUE in the slot to get the cumulative probabilities for all values up to *x*. Place FALSE in the slot to get the value of the probability density function for that combination of *x*, the mean, and the standard deviation. In this chapter, we are more interested in the cumulative probabilities and will enter TRUE most of the time.
- To compute probabilities from an exponential distribution, select **EXPONDIST** from the **Insert Function's Statistical** menu. In the **EXPONDIST** dialog box, there are three lines to which you must respond. On the first line, **X**, enter the value of *x*. On the second line, **Lambda**, enter the value of lambda. The third line, **Cumulative**, requires a logical response of either TRUE or FALSE. Place TRUE in the slot to get the cumulative probabilities for all values up to *x*. Place FALSE in the slot to get the value of the probability density function for that combination of *x* and lambda. In this chapter, we are more interested in the cumulative probabilities and will enter TRUE most of the time.

MINITAB

- Probabilities can be computed using Minitab for many different distributions, including the uniform distribution, the normal distribution, and the exponential distribution.
- To begin uniform distribution probabilities, select **Calc** on the menu bar. Select **Probability Distributions** from the pulldown menu. From the long second pulldown menu, select **Uniform**. From the dialog box, check how you want the probabilities to be calculated from **Probability density**, **Cumulative probability**, or **Inverse probability**. **Probability density** yields the value of the probability density for a particular combination of *a*, *b*, and *x*. **Cumulative probability** produces the cumulative probabilites for values less than or equal to *x*. **Inverse probability** yields the inverse of the cumulative probabilites. Here we are mostly interested in **Cumulative probability**. On the line, **Lower endpoint:**, enter the value of *a*. On the line, **Upper endpoint:**, enter the value of *b*. If you want to compute probabilites for several values of *x*, place them in a column, list the column location in **Input column**. If you want to compute the probability for a particular value of *x*, check **Input constant**, and enter the value of *x*.
- To begin normal distribution probabilites, select **Calc** on the menu bar. Select **Probability Distributions** from the pulldown menu. From the long second pulldown menu, select **Normal**. From the dialog box, check how you want the probabilities to be calculated from **Probability density**, **Cumulative probability**, or **Inverse probability**. **Probability density** yields the value of the probability density for a particular combination of μ, σ and *x*. **Cumulative probability** produces the cumulative probabilities for values less than or equal to *x*. **Inverse probability** yields the inverse of the cumulative probabilities. Here we are mostly interested in **Cumulative probability**. In the space beside **Mean,**enter the value of the mean. In the space beside **Standard deviation**, enter the value of the standard deviation. If you want to compute probabilities for several values of *x*, place them in a column, list the column location in **Input column**. If

you want to compute the probability for a particular value of x, check **Input constant**, and enter the value of x.

- To begin exponential distribution probabilites, select **Calc** on the menu bar. Select **Probability Distributions** from the pulldown menu. From the long second pulldown menu, select **Exponential**. From the dialog box, check how you want the probabilities to be calculated from **Probability density**, **Cumulative probability**, or **Inverse probability**. **Probability density** yields the value of the probability density for a particular combination of x_0 and μ. **Cumulative probability** produces the cumulative probabilities for values less than or equal to x. **Inverse probability** yields the inverse of the cumulative probabilities. Here we are mostly interested in **Cumulative probability**. In the space beside **Scale**, enter a scale value to define the exponential distribution. The scale parameter equals the mean, when the threshold parameter equals 0. *Note:* Minitab uses the mean, $\mu = 1/\lambda$, not the value of λ. In the space beside **Threshold**, enter a threshold number to define the exponential distribution. If you want to compute probabilities for several values of x, place them in a column, list the column location in **Input column**. If you want to compute the probability for a particular value of x, check **Input constant**, and enter the value of x.