

There was a report in *The Wall Street Journal* (January 9, 1990) about Saudi Arabia making another discovery of light, high-value crude oil. Analysts wondered about the effect of this find on international oil prices \triangle

Chapter 5 discussed discrete probability distributions, where random variables assume only specific designated values. In contrast, continuous probability distributions involve random variables that can assume any value within a specified range. These variables are called *continuous random variables*. This chapter examines two special continuous variables of particular importance to managers.

WHY MANAGERS NEED TO KNOW ABOUT CONTINUOUS PROBABILITY DISTRIBUTIONS

Many real-world observations can be modeled using the key distribution discussed in this chapter: the normal distribution. If the actual situation meets the conditions of this important model, answers of value to the decision maker can be found without requiring expensive and time-consuming observations. In addition, the normal distribution is the key to understanding the most important concept in statistics, which will be discussed in the next chapter.

As for the price of oil, this price may fluctuate around a central value, given a steady market. Such fluctuations might very well be modeled with the normal distribution. If so, a series of prices that seemed out of line with this distribution could signal an upward or downward trend.

CONTINUOUS RANDOM VARIABLES

Continuous Random Variable Chapter 5 introduced the concept of a discrete random variable, one that can assume only certain, predetermined values. A variable that can assume any value within some range is called a **continuous random variable**. Such a variable is measured on a numerical scale, that is, constitutes either interval or ratio data.

A continuous random variable is measured on a numerical scale. Each observation of the random variable can assume any value within some specified range.

There are numerous examples of continuous random variables, since most measurements made in business applications are of this type. An automobile's average number of miles per gallon of gasoline is one example. A small car could average somewhere between 25 and 30 miles per gallon, but its average mileage would be unlikely to exactly equal some round number. A claimed mileage of 28 miles per gallon, for example, is commonly understood to mean approximately 28. In fact, it is almost impossible for this car to average precisely 28 miles per gallon, since that implies an average mileage of 28.0000 . . . , out to an infinite number of decimal places. Since there is only one value of interest, 28, out of an infinite number of possible values, the probability is equal to $1/\infty$, or approximately zero.

When dealing with a continuous random variable, one cannot seek the probability that the variable will assume a specific value. Rather, the probability that the variable

SITUAT

THE

Uniform I

HAPTER SIX

CONTINUOUS PROBABILITY DISTRIBUTIONS

167

abia making t of this find

les assume ons involve e variables continuous

scussed in aditions of ad without ormal disics, which

e, given a mal distriion could

m assume thin some tred on a

obserrange.

nost mea-'s average d average would be miles per In fact, it since that f decimal umber of

robability e variable will fall within a specified range is sought. The following two questions illustrate the difference between discrete and continuous random variables:

- 1. What is the probability that the number of people arriving at the checkout counter in the next 30 seconds will be exactly three?
- 2. What is the probability that the lifetime of an electronic switch will be between four years and six years?

The first question suggests the use of a discrete probability distribution. There can only be a discrete number of persons appearing at the checkout counter, such as zero, one, two, etc. In fact, if the arrivals occur randomly and the average arrival rate is known, the answer to this question can be computed using the Poisson distribution, as discussed in Chapter 5.

The second question suggests a continuous probability distribution. Any lifetime is possible within some reasonable range. A value between 4.385 years and 4.587 years is possible, as is a value between 5.7/8 years and 5.15/16 years. The probability that the observed lifetime will fall into a particular range is properly requested, since the probability of the lifetime being precisely equal to some particular value is approximately zero.

SITUATION 6.1

Joe Price, superintendent of the Justus Bag Company, wants to check the weight of grain bags being filled by a machine currently used to bag wheat. Joe knows that the correct weight for each bag is 125 pounds and that there is some room for variability in this weight. He finds from a preliminary sample that the bag weights are evenly distributed between 115 and 135 pounds. The range of tolerance for fill weight for shipping purposes has been determined as ± 5 pounds. Mr. Price wants to know what percent of the bags fall into this tolerance range.

THE UNIFORM DISTRIBUTION

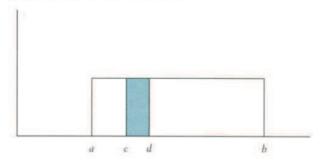
Uniform Distribution

There are specific theoretical probability distributions for continuous random variables, just as there are for discrete random variables. The binomial and Poisson distributions, discussed in Chapter 5, are examples of the latter. The **uniform distribution** is sometimes appropriate when continuous random variables are being observed and where the outcome of an observation is equally likely to occur within any given segment of equal size in the specified range.

The uniform distribution describes a random variable that is as likely to occur in one segment of a given size within a specified range as another.

A uniform distribution is depicted in Figure 6.1. The range within which the random variable can appear is specified by the values a and b. The probability "curve" of Figure 6.1 is of uniform height at all points between a and b, which suggests equal probabilities of the random variable appearing within any segment of a given width in this range. The area of the rectangle is equal to 1.00, or 100%.

FIGURE 6.1 Uniform Distribution



Furthermore, the probability of the variable falling between any two points c and d (see Figure 6.1) is equal to the percentage of the total range included between c and d and can be computed using Equation 6.1.

$$P(c \le X \le d) = \frac{d - c}{b - a} \tag{6.1}$$

It should be noted that since the probability of a value X being precisely equal to any specified value is approximately zero, the equality sign may or may not be included in the specification of a range. That is,

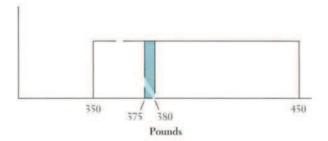
$$P(c < X < d) = P(c \le X \le d)$$

This applies to Equation 6.1 and to any other interval specification for a continuous variable.

EXAMPLE 6.1 Ernie Longnecker, a foreman at Trentwood Aluminum, knows that the weights of cast aluminum blocks are uniformly distributed within a range of 350 to 450 pounds. If Ernie selects a block at random, what is the probability that it weighs between 375 and 380 pounds? Since the random variable, block weight, is known to be uniformly distributed, a particular block weight has the same probability of appearing in any segment of specified width in the range 350 to 450. To answer the question, it is necessary only to compute the percentage of the desired range relative to the total possible range. This example is illustrated in Figure 6.2, and the calculation is:

$$P(c \le X \le d) = \frac{d-c}{b-a} = \frac{380-375}{450-350} = \frac{5}{100} = .05$$

FIGURE 6.2 Uniform Distribution



Mea Dev

CON

HAPTER SIX

CONTINUOUS PROBABILITY DISTRIBUTIONS

Mean and Standard

Deviation of Uniform

169

Five percent of the blocks fall in the interval 375 to 380. In other words, there is a 5% chance that any randomly selected block will have a weight in this interval.

EXAMPLE 6.2 Marilyn Horner knows that the years of company experience for Pacific Western Industries factory workers form a uniform distribution with a minimum of 0 and a maximum of 12.5 years. She wants to select an employee randomly and determine the probability that this person has between 2.5 and 7.4 years of experience with the company. This question can be reworded: What percent of the total range, 0 to 12.5, is the interval 2.5 to 7.4? The answer is:

$$P(c \le X \le d) = \frac{d-c}{b-a} = \frac{7.4-2.5}{12.5-0} = \frac{4.9}{12.5} = .392$$

There is about a 39% chance that an employee's time with Pacific Western Industries falls in the range 2.5 to 7.4 years.

The mean and standard deviation of a uniform distribution are sometimes of interest. The average value of a random variable that follows a uniform distribution falls halfway between the range extremes. In other words, the mean is simply the average of the range boundaries; this calculation is shown in Equation 6.2. The standard deviation of a uniform distribution appears in Equation 6.3.

Mean of uniform distribution =
$$\frac{a+b}{2}$$
 (6.2)

Standard deviation of uniform distribution =
$$\sqrt{\frac{(b-a)^2}{12}}$$
 (6.3)

where a = Lower limit of range

b =Upper limit of range

EXAMPLE 6.3 Tamsen Stone, accountant for The Green Machine High-Performance Yard Care Tools Company, has determined that the ages of accounts receivable are uniformly distributed within a range of 4.7 weeks to 29.4 weeks. Tamsen needs to briefly describe these account ages. The mean and standard deviation of these values provide easily understood summaries of the account age data. These values are calculated using Equations 6.2 and 6.3:

Mean =
$$\frac{4.7 + 29.4}{2}$$
 = 17.05
Standard deviation = $\sqrt{\frac{(29.4 - 4.7)^2}{12}}$ = 7.13

The distribution of account receivable ages has a mean of 17.05 weeks and a standard deviation of 7.13 weeks. Tamsen knows that the average account receivable is 17.05 weeks old and that the ages deviate from their mean by a standard amount of 7.13 weeks.

The MINITAB computer software package can be used to compute uniform probabilities. MINITAB commands for computing the probability for Example 6.2 are:

ints c and een c and

(6.1)

equal to included

ntinuous

nows that ge of 350 it weighs mown to ppearing juestion, the total is:

MTB> SET C1
DATA> 2.5
DATA> 7.4
DATA> END
MTB> CDF C1;
SUBC> UNIFORM 0 12.5
2.5 0.2000
7.4 0.5920
MTB> STOP

The CDF command computes the cumulative probability of a value less than or equal to X. In this case, the data values were set at X=2.5 and X=7.4. The uniform command solved for the probability of an item falling below 2.5 (.2000) and the probability of an item falling below 7.4 (.5920). Therefore, the probability of an item falling between 2.5 and 7.4 is (.5920 - .2000) = .3920.

SITUATION 6.1 Resolved

The grain mill superintendent, Joe Price, wants to know what percent of the filled wheat bags are within 5 pounds of the desired weight, 125 pounds. Since he has observed that weights are equally distributed between 115 and 135 pounds, the uniform distribution can be used as a model. Joe needs to determine what percent of the total range is constituted by the desired range using Equation 6.1:

$$P(c \le X \le d) = \frac{d-c}{b-a} = \frac{130-120}{135-115} = \frac{10}{20} = .50$$

Fifty percent of the bags will fall within 5 pounds of the desired weight of 125 pounds. Joe can now decide whether or not this is an acceptable quality level, and he can either continue to deliver wheat bags at this level or take the time to improve the filling machine performance.

The mean and standard deviation of bag weight might also be relevant to Joe's decision. Using Equations 6.2 and 6.3, these summary values are calculated as follows:

Mean =
$$\frac{115 + 135}{2}$$
 = 125 pounds
Standard deviation = $\sqrt{\frac{(135 - 115)^2}{12}}$ = 5.77 pounds

EXERCISES

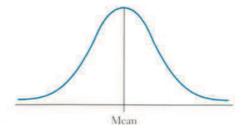
- 1. Explain the difference between the uniform distribution and the Poisson distribution.
- 2. If you were to graph a uniform distribution, what would it look like?
- 3. What continuous random variables are of interest to a credit union studying daily customer transactions?
- 4. Which of the following random variables can be classified as continuous?
 - a. The number of people arriving at a toll gate
 - b. The number of ounces in a bottle of milk
 - c. Tomorrow's temperature
 - d. The number of sales made by a real estate agent this month
 - e. The weight of a steak bought at the supermarket
 - f. The number of offers received on a car for sale
 - g. Number of errors found in a company's accounts receivable records
 - h. Length of time a patient waits for a doctor
 - i How long it took you to travel to school today

- 5. Which of the variables listed in the company data base found in Appendix C are continuous variables?
- 6. The manager of air operations for a West Coast commuter airline has just read an article in *The Wall Street Journal* (November 6, 1989) indicating that air service between the United States and Japan is to be greatly expanded. She is concerned about the flight time between San Diego and Los Angeles, as that route is a key link to the orient flights. The flight time between Los Angeles and San Diego is known to be uniformly distributed between 80 and 100 minutes.
 - a. What is the average flight time between the two cities?
 - b. What is the standard variation in flight time between the two cities?
 - c. What percentage of flights can be expected to take between 85 and 95 minutes?
 - d. What is the probability that a flight will take more than 96 minutes?
 - e. What is the probability that a flight will take less than 80 minutes?
- 7. General Equipment Manufacturing believes that one of its rolling machines is producing sheets of aluminum of varying thickness. The machine typically produces sheets between 75 and 150 millimeters thick. It is known that this random variable is uniformly distributed. Sheets less than 100 millimeters thick are unacceptable to buyers and must be scrapped.
 - a. What is the average thickness of the aluminum sheets produced by this machine?
 - b. What is the standard variation in thickness of the aluminum sheets produced by this machine?
 - c. What is the probability that a sheet produced by this machine will have to be scrapped?
- 8. The number of bricks a mason will lay in one hour is uniformly distributed with a mean of 200. The difference between the smallest and the largest number of bricks the mason can lay is 100. What is the probability that the mason will lay between 160 and 190 bricks?
- A traffic light is set to switch from red to green according to a uniform distribution with a mean of 45 seconds. The difference between the smallest and the largest number of seconds it takes the light to switch is 8 seconds.
 - a. Compute the standard deviation of this distribution.
 - b. What is the probability that the light will take at least 43 seconds to switch?
 - c. What is the probability that the light will take less than 43 seconds to switch?

NORMAL DISTRIBUTION

In the 18th century, astronomers observed that repeated measurements of the same value (such as the mass of an object) tended to vary. When a large number of such observations were recorded, organized into a frequency distribution, and plotted on a graph, the recurring result was the shape shown in Figure 6.3. It was subsequently discovered that this distribution could be closely approximated with a continuous distribution sometimes referred to as the *normal distribution*. This distribution is also referred to as the bell-shaped curve or the Gaussian curve.

FIGURE 6.3 Normal Distribution



less than or = 7.4. The (.2000) and ability of an

of the filled ince he has the uniform of the total

ight of 125 level, and to improve

ant to Joe's las follows:

ibution.

ily customer

There are three reasons why the normal distribution is the most important theoretical distribution in business statistics:

- The normal distribution approximates the observed frequency distributions of many natural and physical measurements, such as IQs, weights, heights, sales, product lifetimes, and variability of human and machine outputs.
- 2. The normal distribution can often be used to estimate binomial probabilities when *n* is greater than 20.
- The normal distribution is a good approximation of distributions of both sample means and sample proportions of large samples (n ≥ 30). This concept will be examined in depth in Chapter 7.

Many variables in nature, including the field of business, have numerical observations that tend to cluster around their mean. In other words, it is more likely that an observation will be close to the mean of the data collection than far away. When this condition holds, the theoretical **normal distribution** may provide a good model of the data collection.

Normal Distribution

The normal distribution is a continuous distribution that has a bell shape and is determined by its mean and standard deviation.

Figure 6.3 shows the shape of a typical normal curve. Note that the curve is symmetrical: half the area of the curve lies above the mean and half below it, and these halves are mirror images. Also note that as one moves away from the mean toward either of the tails, the height of the curve decreases. This corresponds to a decreasing probability of finding a value the further one moves from the mean.

The theoretical normal distribution approaches infinity at the positive and negative ends of the curve. In practice this aspect is ignored since the probability of an observation becomes very small as the distance from the mean of the curve increases. A discussion of the mathematical equation for the normal curve is beyond the scope of this text, but the formula is presented here to show its dependence on two key parameters, the mean and standard deviation of the distribution. The so-called probability density function for the normal distribution is shown in Equation 6.4.

$$f(X) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(1/2) ((X - \mu)/\sigma)^2}$$
(6.4)

where X = Any value of the continuous random variable

 μ = Mean of the normal random variable

 σ = Standard deviation of the normal random variable

e = 2.71828 . . . (natural log base)

 $\pi = 3.1416...$ (used to find circumference of circle)

Mathematical computations utilizing Equation 6.4 are beyond the scope of this text. However, note that there are two unspecified values in the equation: the mean of the distribution (μ) and the standard deviation (σ) . Both e and π are constants with known values. If the mean and standard deviation are known, then a specific normal curve has been identified, and probabilities can be found.

Figure 6.4 shows three normal curves, each with a different mean. However, the

HAPTER SIX

theoretical

tions of its, sales,

bilities

th sample t will be

ical obserlikely that ay. When ood model

ipe and

e curve is w it, and the mean onds to a lean.

I negative an obsercreases. A e scope of ey paramrobability

(6.4)

he of this he mean ants with c normal

ever, the

CONTINUOUS PROBABILITY DISTRIBUTIONS

three distributions have the same standard deviation. This can be seen by the identical shape or dispersion of each curve.

FIGURE 6.4 Normal Curves with Different Means but the Same Standard Deviation

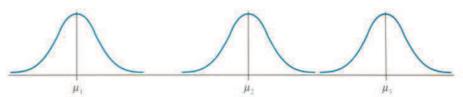


Figure 6.5 shows three normal curves, all with the same mean. However, the extent of their variability is different. Curve A has a larger standard deviation than curve B, and curve C has the smallest standard deviation of all.

Figures 6.4 and 6.5 illustrate the point that both the mean and the standard deviation must be known before a specific normal curve can be identified out of the infinite number of curves available.

Characteristics of the normal distribution

- 1. The normal distribution has a bell shape and is symmetrical about its mean.
- Knowledge of the mean and standard deviation is necessary to identify a specific normal distribution.
- 3. Each combination of mean and standard deviation specify a unique normal distribution
- The normal distribution extends indefinitely in either direction from the mean.
- 5. The normal distribution is measured on a continuous scale, and the probability of obtaining a precise value is zero.
- 6. The total area under the curve is equal to 1.0, or 100%. Fifty percent of the area is above the mean, and 50% is below the mean.
- 7. The probability that a random variable will have a value between any two points is equal to the area under the curve between those points. This area can be determined using either calculus or the standard normal table in Appendix E.

EXERCISES

- 10. Explain the difference between the uniform distribution and the normal distribution.
- 11. Explain the difference between the binomial distribution and the normal distribution.
- 12. If you were to graph a normal distribution, what would it look like?
- 13. Which of the following are characteristics of the normal distribution?
 - a. The normal distribution is skewed.
 - Knowledge of the median and standard deviation is necessary to construct a specific normal distribution.
 - e. Each combination of mean and standard deviation defines a unique normal distri-

- d. The normal distribution extends indefinitely in either direction from the mean.
- e. The normal distribution is measured on a discrete scale.
- f. The total area under the curve is equal to 1.0 or 100%.
- g. The probability that a random variable will have a value between any two points is equal to the area under the curve between those points.
- 14. Why is the normal distribution the most important theoretical distribution in business statistics?

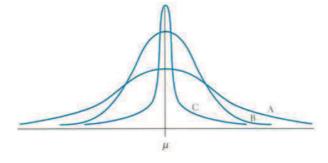
SITUATION 6.2

Applied Business Systems, Inc., a mail-order house, has been studying the distances involved in shipping orders to its customers. After collecting a sample of shipping distances over the past two months and plotting them on a graph, Julie Blocker, the company analyst, decides that these distances greatly resemble a normal distribution. The mean distance shipped is 125 miles, with a standard deviation of 28 miles. There is a shipping cost increase at 135 miles, and Julie feels that management should know what percent of orders exceed this distance. Since the company's own drivers can deliver any order within 100 miles, she feels that management should also know what percent of orders fall into this category.

FINDING NORMAL CURVE AREAS

To find normal curve areas, it is necessary to use integral calculus along with Equation 6.4, and the computations are complex. For this reason, it would be convenient if a table for normal curve areas were available, as they are for the binomial and Poisson distributions. Unfortunately, more than one table would be necessary. Figures 6.4 and 6.5 illustrate the point that the normal distribution is actually an infinitely large family of distributions, one for each possible combination of mean and standard deviation. The number of normal distributions is therefore unlimited, and providing a table of probabilities for each combination of mean and standard deviation would not be practical.

FIGURE 6.5 Normal Curves with the Same Mean but Different Standard Deviations



Fortunately, the problem of dealing with an infinite family of normal distributions can be solved by transforming all normal distributions to the **standard normal distribution**, which has a mean equal to 0 and a standard deviation equal to 1.

CONTINUOUS PROBABILITY DISTRIBUTIONS

z Value

Normal Curve Areas

175

The **standard normal distribution** has a mean equal to 0 and a standard deviation equal to 1.

Any normal distribution can be converted to the standard normal distribution by standardizing each of its observations in terms of *z* values. The *z* value measures the distance in standard deviations between the mean of the normal curve and the *X* value of interest. In Chapter 4, the *z* value was called the standard score.

The z value is a measure of the number of standard deviations between the mean or center of a normal curve and the value of interest.

Equation 6.5 shows how the observations from any normal distribution can be standardized by forming a standard normal distribution of z values.

$$z = \frac{X - \mu}{\sigma} \tag{6.5}$$

where z = Number of standard deviations from mean

X = Value of interest

 μ = Mean of distribution

 σ = Standard deviation of distribution

Examples 6.4 and 6.5 illustrate applications of Equation 6.5.

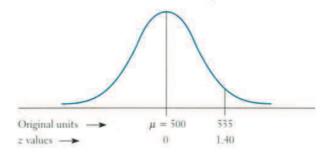
EXAMPLE 6.4 A normal curve has a mean of 500 and a standard deviation of 25. An analyst is interested in the value 535 and wants to find the normal curve areas both above and below this value.

What is the correct z value for 535? From Equation 6.5, the z value is:

$$z = \frac{X - \mu}{\sigma} = \frac{(535 - 500)}{25} = 1.40$$

Figure 6.6 shows both the actual scale of this normal distribution and the standardized scale for the standard normal distribution. Note that the standard normal distribution is created by converting the actual values to z values.

FIGURE 6.6 Normal Curve for Example 6.4



CHAPTER SIX

mean.

wo points is

in business

e distances of shipping locker, the istribution. iles. There ould know drivers can

know what

h Equation venient if a nd Poisson figures 6.4 nitely large

d standard

providing

tion would

tions

stributions mal distriEXAMPLE 6.5 Coffee prices recently tumbled to a 14-year low of 76.5 cents a pound, according to *The Wall Street Journal* (October 4, 1989). Suppose a small importer in Florida watches these prices closely, as they are directly related to profits.

The importer averages 2,700 pounds per month with a standard deviation of 130 pounds. What is the z value for 2,200 pounds on this distribution?

$$z = \frac{2200 - 2700}{130} = -3.85$$

After the z value for the standard normal distribution has been calculated, the next step is to use the standard normal table to look up the corresponding area under the curve. The standard normal table, located in Appendix E, is designed to be read in units of z, the number of standard deviations from the mean. The table shows the area under the curve between the mean and selected values of z. It is important to remember that this table specifies the area between the *center* of the curve and the point specified by the z value.

Since the normal distribution is symmetrical about its mean, the left half of the curve is a mirror image of the right half. Because of this symmetry, the standard normal table provides only the right half (i.e., positive z values) of the distribution. If the z value of interest is negative, the minus sign is ignored, and the area is obtained in the same manner as if the z value were positive. For example, the area under the curve between the mean and +2 standard deviations is exactly equal to the area under the curve between the mean and -2 standard deviations.

The reason for the empirical rule presented in Chapter 4 can now be explained. Using the standard normal table, it is possible to find the percentages of items on a standard normal distribution that will fall within 1, 2, and 3 standard deviations of the mean. These values are calculated in Example 6.6.

EXAMPLE 6.6 What percent of the values on a normal curve can be expected to fall within 1, 2, and 3 standard deviations of the mean? The standard normal table is entered with z = 1, z = 2, and z = 3, and the areas are noted. These areas are then doubled to account for values both above and below the mean. These areas are

$$z = 1$$
: (.3413) (2) = .6826
 $z = 2$: (.4772) (2) = .9544
 $z = 3$: (.4987) (2) = .9974

These values establish the empirical rule that was presented in Chapter 4: about 68% of the values are within 1 standard deviation of the mean, about 95.5% are within 2 standard deviations, and over 99.7% are within 3 standard deviations. Figure 6.7 shows these areas.

Examples 6.7 through 6.13 illustrate applications of the standard normal table.

EXAMPLE 6.7 A normal distribution has a mean of 5 and a standard deviation of 2.1. What is the probability that a numerical value randomly drawn from this distribution will lie between 5 and 7? In this example, the question is phrased properly because the normal distribution represents a continuous random variable, and probabilities must be stated for a range, not a specific value. Figure 6.8 shows the desired area. Using Equation 6.5, the z value is: z = (7 - 5)/2.1 = 0.95.

Since the area desired extends from the mean of the curve (5) to a point of interest, X (7), the desired area can be read directly from the standard normal table using the



d, the next under the be read in shows the portant to ve and the

half of the e standard stribution. is obtained under the area under

explained. tems on a viations of

xpected to mal table areas are areas are:

r 4: about are within igure 6.7

il table.

viation of this distriproperly and probne desired

f interest. using the



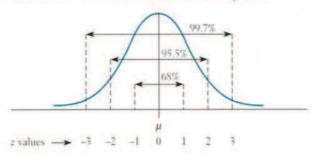
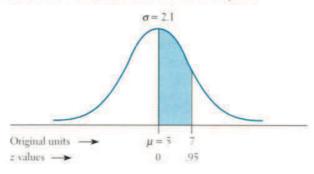


FIGURE 6.8 Normal Curve Area for Example 6.7



calculated z value. Note the graph at the top of the table in Appendix E, which suggests that the areas under the curve are measured from the center of the curve to the point of interest.

The left column of the standard normal table lists the z value with only one digit on each side of the decimal point. Since the z value for this example is 0.95, the 0.9 row is used. The values across the top of the table give the second digit to the right of the decimal point. Since this digit is 5 in this example, the .05 column is used. At the intersection of the 0.9 row and the .05 column of the table is the value .3289. This is the percentage of the total area under the curve found between the center of the distribution and a point 0.95 standard deviations away. There is about a 33% chance of drawing a value from this normal distribution that lies between 5 and 7.

EXAMPLE 6.8 A normal distribution with a mean of 250 and a standard deviation of 12 is being studied. It represents the number of mutual fund shares traded on a local exchange. The fund managers note that mutual fund sales nationally rose 4.2 billion in August 1989 to the highest value in two years (The Wall Street Journal, October 2, 1989). What is the probability that a share value randomly selected from this distribution will fall between 225 and 250? The z value is: z = (225 - 250)/ 12 = -2.08

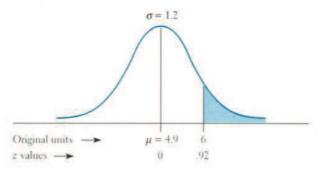
The negative sign of the z value indicates that the X value of interest (225) is less than the mean of the distribution (250); the negative sign is ignored when the z value is looked up in the standard normal table. As in Example 6.7, since this area is measured from the mean of the curve to some point, the desired area or probability can be read directly from the table. The 2.0 row of the table is used along with the .08 column, and the desired area is read from the table as .4812. Thus, there is about a 48% chance that a randomly selected item from this distribution will fall in the interval 225 to 250.

Examples 6.7 and 6.8 both involved areas measured from the mean or center of the normal curve to a point of interest. As a result, the normal curve area could be read directly from the table. There are three other kinds of normal curve areas that might be of interest:

- 1. An area that lies at one end of the curve (see Example 6.9).
- An area that spans the curve mean; that is, some of the area is above the mean, and some is below (see Example 6.10).
- 3. An area that is either above the mean or below it but does not include the mean (see Example 6.11).

EXAMPLE 6.9 A normal distribution has a mean of 4.9 and a standard deviation of 1.2. What percent of the area under this curve lies above 6? This question can be rephrased: What is the probability that an item randomly selected from this distribution will have a value above 6? Figure 6.9 shows the desired area.

FIGURE 6.9 Normal Curve Area for Example 6.9



To solve this problem, it is first necessary to find the area from the mean of the curve (4.9) to the X value of interest (6). This area is found by calculating the z value and referring to the standard normal table. The z value is: z = (6 - 4.9)/1.2 = 0.92.

The area from the standard normal table corresponding to a z value of 0.92 is .3212 (found at the intersection of the 0.9 row and the .02 column). To find the desired area, it must be remembered that exactly 50% of a normal curve lies above the mean and 50% lies below it. If half the normal curve area lies above the mean of 4.9, and if the area .3212 lies between 4.9 and 6, then the remaining portion of the half must lie above 6. Therefore, the desired area is: .5000 - .3212 = .1788. There is about an 18% chance that a randomly selected item will have a value greater than 6 since about 18% of the curve area lies above this value.

Incidentally, the area in one tail of the normal curve is of particular importance in many statistical applications. In the chapters on statistical inference, this area will be referred to as the *probability value*, or *p*-value, and will be the basis on which a conclusion about the population will be reached. In subsequent chapters, references to *p*-value will simply refer to the area in either the upper or lower end of a curve.

CHAPTER SIX

fall in the

or center of ea could be a areas that

the mean,

e the

l deviation ion can be listribution

can of the z value = 0.92. of 0.92 is o find the dies above the mean portion of

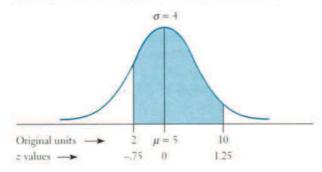
= .1788. ue greater

area will on which ers, referend of a CONTINUOUS PROBABILITY DISTRIBUTIONS

....

EXAMPLE 6.10 What is the probability that an item drawn from an inventory of television tubes will have a shelf life between 2 and 10 weeks? The ages of these parts are known to follow a normal distribution with a mean of 5 weeks and a standard deviation of 4 weeks. Figure 6.10 shows the desired area.

FIGURE 6.10 Normal Curve Area for Example 6.10



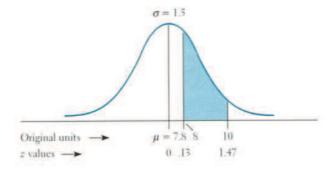
This problem actually involves two normal curve areas, each measured from the center of the curve, that must be determined separately. The two areas will be added together to find the desired value. The two z values and their areas are:

$$z = (2 - 5)/4 = -0.75 \rightarrow .2734$$
 (Area between 2 and 5)
 $z = (10 - 5)/4 = 1.25 \rightarrow .3944$ (Area between 5 and 10)
Sum = .6678 (Area between 2 and 10)

There is about a 67% chance that a randomly selected television tube will have a lifetime of between 2 and 10 weeks.

EXAMPLE 6.11 Years of company experience of a pool of first-line supervisors at Future Business Products, Inc., are known to follow a normal distribution with a mean of 7.8 years and a standard deviation of 1.5 years. This information came to light after the president of the company began an investigation into the company's retention problem. This investigation was started after the president read an article (*Inc.* magazine, May 1989, p. 132) about the advantages of bonus and incentive systems for employees. What is the probability that a randomly selected supervisor has between 8 and 10 years of experience? Figure 6.11 shows the area to be determined.

FIGURE 6.11 Normal Curve Area for Example 6.11



To solve this problem, it is first necessary to find the area between the mean of the curve, 7.8, and 10, the upper value of interest. The area between 7.8 and 8 will then be subtracted from it. The result is the area lying between 8 and 10. The two z values and their areas from the standard normal table are:

$$z = (10 - 7.8)/1.5 = 1.47 \rightarrow .4292$$
 (Area between 7.8 and 10)
 $z = (8 - 7.8)/1.5 = 0.13 \rightarrow .0517$ (Area between 7.8 and 8)
Difference = .3775 (Area between 8 and 10)

There is about a 38% chance that a randomly selected supervisor will have between 8 and 10 years of experience with Future Business Products, Inc.

Sometimes an analyst knows the area under the curve and must find an actual value or a z value. Example 6.12 illustrates this situation.

EXAMPLE 6.12 The Alpine Tire Company has determined from road tests that the mean mileage of its main product is 50,000 miles with a standard deviation of 5,000 miles, and that the collected data are normally distributed. Alpine wishes to offer a warranty that will provide a free replacement for any new tire that fails before the guaranteed mileage. If Alpine wishes to replace no more than 10% of the tires, what should the guaranteed mileage be?

FIGURE 6.12 Normal Curve Area for Example 6.12

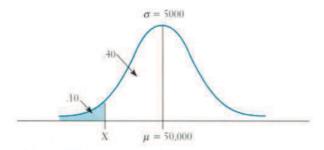


Figure 6.12 shows the area of interest (shaded). To solve this problem, reverse the procedure for using the standard normal table. That is, instead of looking up the area for a particular z value, the analyst looks for the z value that coincides with the area of interest under the curve. This area equals 10%, but to find the appropriate z value in the table, the analyst must subtract 10% from 50% (see Figure 6.12). The area that is closest to .4000 is .3997, which is found at the intersection of the 1.2 row and the .08 column. The z value is thus -1.28 (the negative sign indicates that the z value is below the mean). To find the appropriate mileage for the guarantee, a z value of -1.28 is substituted into Equation 6.5:

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

$$-1.28 = \frac{X - 50,000}{5,000}$$

$$X - 50,000 = -1.28(5,000)$$

$$X - 50,000 = -6,400$$

$$X = 43,600$$

IAPTER SIX

ean of the

will then o z values

CONTINUOUS PROBABILITY DISTRIBUTIONS

Alpine should set their guarantee at 43,600 miles since only 10% of actual tire mileages will fall below this value.

As implied by the preceding example, Equation 6.5 can be manipulated so that X can be solved for directly: $X = \mu + z\sigma$.

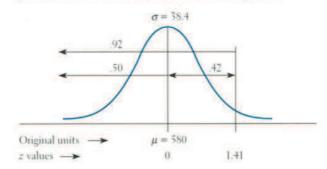
EXAMPLE 6.13 The scores on a national examination follow a normal distribution with a mean of 580 and a standard deviation of 38.4. If a person scores higher than 92% of all those who have ever taken the exam, what score did this person get? With 92% of the scores below this X value, 42% of the scores must lie between the mean of 580 and X (see Figure 6.13). In the standard normal table, an area of .4207 is closest to the desired value of .42. The z value for this area is 1.41, so X is 1.41 standard deviations above the mean:

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

 $X = 580 + (1.41)(38.4) = 634$

Ninety-two percent of the people scored below 634 on the national examination.

FIGURE 6.13 Normal Curve Area for Example 6.13



The MINITAB computer software package can be used to compute normal curve probabilities. The MINITAB commands for computing the normal probability for Example 6.8 are:

MTB> SET C1 DATA> 225 DATA> 250 DATA> END MTB> CDF C1; SUBC> NORMAL 250 12. 225 0.0186 250 0.5000 MTB> STOP

The CDF command computes the cumulative probability of a value less than or equal to X. In this case, the data values were set at X=225 and X=250. The uniform command solved for the probability of an item falling below 250 (.0186) and the probability of an item falling below 250 (.5000). Therefore, the probability of an item falling between 225 and 250 is (.5000 - .0186) = .4814.

between an actual

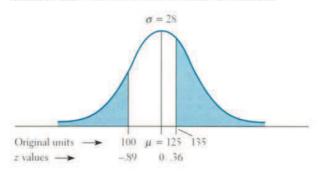
tests that riation of wishes to ils before the tires,

verse the the area the area z value area that and the z value of

SITUATION 6.2 Resolved

Applied Business Systems, Inc., wants to know what percent of its orders are shipped more than 135 miles and what percent are shipped less than 100 miles. From a graph of shipping distances, it is apparent that a normal curve can be used to approximate shipping distances. The mean distance is calculated to be 125 miles with a standard deviation of 28 miles. Figure 6.14 shows the desired areas.

FIGURE 6.14 Normal Curve Areas for Situation 6.2



The z values for the two areas can now be found. Since the areas are on the ends of the curve, the values from the standard normal table must be subtracted from .50 to get the desired results. The calculations are:

$$z = (135 - 125)/28 = 0.36 \rightarrow .1406$$
 (Area between 125 and 135)
 $.5000 - .1406 = .3594$ (Area above 135)
 $z = (100 - 125)/28 = -0.89 \rightarrow .3133$ (Area between 100 and 125)
 $.5000 - .3133 = .1867$ (Area below 100)

About 36% of the orders are shipped in excess of 135 miles. This is a significant percentage, and management should investigate ways of reducing the higher shipping charges for these distances. About 19% of the firm's orders are shipped less than 100 miles. The company should evaluate the cost benefits of using company drivers for this rather low percentage of orders.

Chebyshev's Theorem

Example 6.6 provided areas for the empirical rule, which applies only if the distribution in question follows a normal curve or can be closely approximated with one. For distributions that do not follow a normal curve, areas can be estimated through the use of **Chebyshev's theorem**, which was briefly mentioned in Chapter 4. The theorem states that regardless of how the data values are distributed, at least a certain percentage will lie within a specified range around the mean.

Chebyshev's theorem

Regardless of how the data values are distributed, at least $(1 - 1/k^2)$ of the values will lie within k standard deviations of the mean (for $k \ge 1$).

CONTINUOUS PROBABILITY DISTRIBUTIONS

183

EXAMPLE 6.14 The distribution of ages in a factory do not follow a normal distribution but have a mean of 44.8 with a standard deviation of 9.7. Using Chebyshev's theorem, the percentage of ages within 2 standard deviations of the mean can be calculated:

$$(1 - 1/k^2) = 1 - 1/2^2 = 1 - 1/4 = .75$$

Thus, at least 75% of the ages in the factory are within 2 standard deviations of the mean. In other words, the ages of at least 75% of the employees range from 44.8 - (2)(9.7) to 44.8 + (2)(9.7), or from 25.4 to 64.2 years.

EXERCISES

- 15. In using the standard normal table to look up a z value, what point is used consistently as the reference point?
- 16. Why doesn't the standard normal table contain areas above .5000?
- 17. Why are X values converted to z values?
- 18. Draw a normal curve and shade the area under the curve for each of the following:
 - a. The area between the mean and z = 0.75
 - b. The area between the mean and z = -1.75
 - c. The area between z = -1.25 and z = 0.25
 - d. The area between z = 0.30 and z = 2.00
 - e. The area above z = 1.00
 - f. The area below z = -3.00
- Find the probability of an item falling into each of the shaded intervals drawn in Exercise 18.
- 20. Find the z value that corresponds to each of the areas described:
 - a. Seventy percent of the items lie above this z value.
 - b. Twenty percent of the items lie below this z value.
 - c. Ten percent of the items lie above this z value.
 - d. Sixty percent of the items lie above this z value.
- 21. Home loan applications are normally distributed with mean of \$100,000 and a standard deviation of \$25,000 at the Rainier State Bank. Bank policy requires that applications involving loan amounts in the lowest 5% be submitted to a vice president in charge of authorizing low-income housing loans. What size home loan requests are submitted to this vice president?
- 22. Street lights used in the residential areas served by Utah Power & Electric are constructed to have a mean lifetime of 500 days with a standard deviation of 50 days. If it can be assumed that street light lifetimes are normally distributed, what percentage of the lamps:
 - a. Last longer than two years?
 - b. Last between 400 and 625 days?
 - c. Last longer than 600 days?
 - d. Last between 350 and 450 days?
- 23. The fuel consumption of a fleet of 1,000 trucks is normally distributed with a mean of 12 miles per gallon and a standard deviation of 2 miles per gallon.
 - a. How many trucks will average 11 miles or more per gallon?
 - b. How many trucks will average less than 10 miles per gallon?
 - c. How many trucks will average between 9.5 and 14 miles per gallon?
 - d. How many trucks will average between 9 and 11 miles per gallon?
 - e. What is the probability that a truck selected at random will average over 13.5 miles per gallon?
 - f. Seventy percent of the trucks average more than what mileage?
 - g. Ten percent of the trucks average less than what mileage?

he ends

APTER SIX

shipped

i a graph roximate standard

135)

125)

hipping han 100 wers for

the disth one. hrough 4. The certain

lues

Bino

Con

- 24. Horace Gainey, analyst for the Splice Soft Drink Corporation, is asked to set the fill level for the company's new soft drink dispensers. If the ounces of fill are normally distributed with a mean of μ and a standard deviation of 0.2 ounces, what should Horace set μ at so that 8-ounce cups will overflow only one time in a thousand?
- 25. The Brooks Publishing Company has discovered that the number of words contained in new manuscripts in excess of the number specified in the author's contract is normally distributed with a mean equal to 40,000 words and a standard deviation of 10,000 words. If Brooks wants to be 90% certain that a new manuscript will be less than 250,000 words, how many words should be specified in the author's contract?
- The hourly wage earned by workers at a state correctional institution is normally distributed with a mean of \$9.10 and a standard deviation of \$0.90.
 - a. What is the probability that a worker's hourly wage exceeds \$10,00?
 - b. What is the probability that a worker's hourly rate is between \$7.70 and \$11.50?
 - c. What proportion of the workers earn less than \$8.50?
 - d. Half the workers earn more than what amount?

SITUATION 6.3

Professional Marketers, an independent research firm, has been asked to take a random sample of 1,000 voters from Lincoln County to determine whether or not they approve of a proposed bond issue. At least 60% of the voters on election day must approve the issue for it to pass. Shirly Knight, president of Metro Growth, the organization sponsoring the bond issue, is assuming that there is 60% support among all voters. She wants to know the probability that the sample will reveal a support level less than 58% if her assumption of at least 60% favorable is actually true.

NORMAL APPROXIMATION TO THE BINOMIAL

Chapter 5 presented the binomial distribution and various methods of solving binomial problems. As shown in Figure 5.7, if the number of trials (n) is 20 or less, the correct binomial probabilities can be found in the binomial tables of Appendix E. If n is greater than 20 and the probability of success, p, is either very small or very large, the Poisson distribution can provide a good approximation using the method discussed in Chapter 5.

Suppose a binomial problem involves an *n* value larger than 20, so that the binomial tables cannot be used, and a proportion that is neither very small nor very large. In such a case, the Poisson approximation is not appropriate, and another method of solving the problem must be found.

As suggested by Figure 5.7, the normal curve provides a good approximation to the binomial distribution under these conditions. The approximation is made that the discrete binomial distribution follows the continuous normal distribution. This approximation introduces an error into the probability calculations, but this error has been found to be quite small.

When the normal approximation is used, the mean and standard deviation of the binomial distribution must be found so that they can be used to compute normal curve probabilities.

ne fill level distributed set μ at so

ntained in normally 000 words. 000 words,

distributed

1.50?

random approve prove the on sponters. She han 58%

oinomial e correct . If n is ry large, liscussed

arge. In ethod of

ation to that the This aprror has

n of the

Normal Approximation to Binomial

Continuity Correction Factor

The normal distribution provides a close approximation to the binomial distribution when n is large and p is close to .5. As a rule of thumb, this approximation is appropriate when np > 5 and n(1 - p) > 5.

Both the Poisson and binomial distributions are discrete, whereas the normal curve is continuous. Since discrete variables involve only specified values, intervals will have to be assigned to the continuous normal distribution to represent binomial values. For example, continuous values in the range 4.5 to 5.5 correspond to the discrete value 5. This addition and subtraction of .5 to the X value is commonly referred to as the continuity correction factor.

To find the binomial probability of exactly 5 successes, the normal curve approximation is used based on the probability (i.e., the area under the normal curve) between 4.5 and 5.5. If the desired binomial probability is for 15 or more successes, the normal curve area from 14.5 to ∞ is found. If the binomial probability of interest involves 4 or fewer successes, the appropriate normal curve area is from 4.5 to $-\infty$.

Since the number of trials, or sample size, is usually large in practical situations where the normal distribution is used to approximate the binomial, the continuity correction factor typically has a negligible effect. However, this correction is made in the examples and exercises that follow.

FIGURE 6.15 Normal Approximation to the Binomial

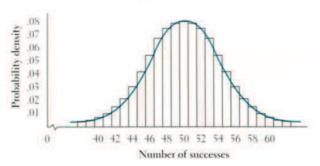


Figure 6.15 shows how the normal curve can be used to give binomial probability approximations. The bars represent binomial probabilities for the distribution of X successes in 100 trials where the probability of success is .5. The normal curve superimposed over the binomial histogram has a mean and standard deviation equal to:

$$\mu = np = (100)(.50) = 50$$

 $\sigma = \sqrt{np(1 - p)} = \sqrt{(100)(.50)(.50)} = 5$

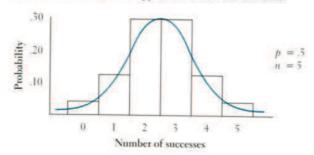
Figure 6.15 shows that the area in any set of bars (the binomial probability) is approximately equal to the corresponding area under the curve (the normal probability).

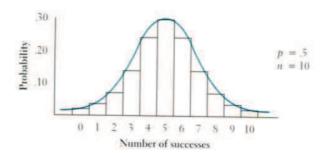
Figure 6.16 also shows how the binomial distribution can be approximated by the normal distribution. In Figure 6.16 three binomial probability distributions have been

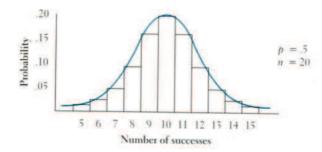
CONTI

plotted for p = .5: n = 5, n = 10, and n = 20. Normal distributions with means of np and standard deviations of $\sqrt{np(1-p)}$ have been superimposed on the binomial distributions. Figure 6.16 indicates that as n increases, the normal distribution more closely approximates the binomial distribution.

FIGURE 6.16 Normal Approximation to the Binomial







EXAMPLE 6.15 The librarian of the John Kennedy Library, Don Lake, randomly selects 100 books from the shelves and inspects them to determine physical condition. In his last inspection, 40% of the books were found to be in an unsatisfactory condition. Assuming this percentage still holds, what is the probability that Don will find at least 50% of the sampled books unsatisfactory?

This is a binomial situation, since a book is either in good condition or not and since independence between selections is ensured due to random sampling. The binomial problem can be stated:

$$n = 100$$
 $p = .40$ $P(X \ge 50) = ?$

SITI

Resol

HAPTER SIX

ith means

binomial

tion more

CONTINUOUS PROBABILITY DISTRIBUTIONS

187

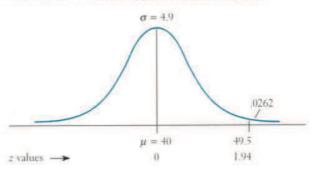
The normal curve can provide an approximation to the correct binomial answer, since n is too large for the binomial tables and the proportion (.40) precludes use of the Poisson approximation (np = 40). The mean and standard deviation of the binomial situation are first found:

$$\mu = np = (100)(.40) = 40$$

 $\sigma = \sqrt{np(1-p)} = \sqrt{(100)(.40)(.60)} = 4.9$

Figure 6.17 shows the area that needs to be computed from the standard normal table. Note that the continuity correction factor is used. The desired area to be found in the standard normal table is from 40 to 49.5. The z value for the desired area is: z = (49.5 - 40)/4.9 = 1.94.

FIGURE 6.17 Normal Approximation to Binomial



The standard normal table indicates that the probability of obtaining a z value of 1.94 is extremely high. Therefore, Don can assume that an area of 47.38% is being covered by an X value 1.94 standard deviations above the mean. There is only a small chance, 2.62%, that 50 or more of the sampled books will be unsatisfactory.

SITUATION 6.3 Resolved Joan Hernandez of Professional Marketers is assigned the task of answering Shirly Knight's question. She formulates the situation as a binomial problem:

$$n = 1,000$$
 $p = .60$ $P(X < 580) = ?$

The normal approximation will be used for this problem. The mean and standard deviation for the binomial situation are found, and the z value is computed:

$$\mu = np = (1,000)(.60) = 600$$
 $\sigma = \sqrt{np(1-p)} = \sqrt{(1,000)(.60)(.40)} = 15.5$
 $z = (579.5 - 600)/15.5 = -1.32$

From the standard normal table, the area for a z value equal to -1.32 is .4066. Joan must subtract this area from .5000 to find the area in the lower end of the normal curve. The correct area is: .5000 - .4066 = .0934.

Joan explains to Shirly that there is about a 9% chance that the sample will result in less than 580 favorable responses if, in fact, the voter population has a favorable response rate of 60%.

ndomly ndition. ndition. at least

not and g. The

EXERCISES

- 27. When is the Poisson distribution a good approximation for the binomial distribution?
- 28. When is the normal distribution a good approximation for the binomial distribution?
- 29. What problem occurs when the normal distribution is used to approximate the binomial distribution? (Hint: This problem does not occur when the Poisson distribution is used.)
- 30. Explain the concept behind the continuity correction factor.
- 31. For a binomial distribution with n = 50 and p = .40, find:
 - a. $P(X \ge 20)$
 - b. $P(X \le 15)$
 - c. $P(X \ge 17)$
 - d. $P(X \le 24)$
- 32. In the past, Sue Megard has made mistakes on 10% of the tax returns she prepared. What is the probability that she will make mistakes on less than 5 of the 70 returns she prepares for 1990? Assume statistical independence.
- 33. A salesperson has a 30% chance of making a sale to any customer called upon. If 110 calls are made during a particular month, what is the probability that:
 - a. At least 30 sales are made?
 - b. Less than 40 sales are made?
- 34. An automatic machine produces washers, 4% of which are defective. If a sample of 300 washers is drawn at random from the production of this machine, what is the probability of observing at least 10 defectives?
- 35. As of 1987, according to the Statistical Abstract of the United States: 1989, 23.6% of the country's 89.5 million households were classified as one-person households. If 1,500 households are randomly selected to participate in an Arbitron survey to determine television ratings, find the probability that no more than 375 of these are one-person households.
- 36. Management of the Boulevard Restaurant has been studying whether or not new customers return within a month. The collected data reveal that 60% of the new customers have returned. If 90 new customers dine at the Boulevard Restaurant this month, what is the probability that at least 50 will return next month?

SUMMARY

This chapter has presented two continuous probability distributions that are useful models of many practical business situations: the uniform distribution and the normal distribution. The normal curve in particular is used frequently in business applications because so many business variables have values corresponding to the two key attributes of the normal curve: the values are symmetrically clustered around their mean, and there are more values close to the mean than far from it.

It is important to understand how to compute normal curve probabilities given the mean and standard deviation of the distribution because of the many situations that can be modeled with the normal curve. An even more important reason for understanding normal curve areas is discussed in the next chapter. The normal curve underlies the inferential process that is the basis of all statistics: the process of reaching conclusions about populations based on sample evidence.

Finally, the normal approximation to the binomial distribution was presented. Along with the binomial procedures of Chapter 5, four methods of solving a binomial problem have been presented: the binomial formula, the binomial tables, the Poisson approximation, and the normal approximation.

ibution?

ne binomial

on is used.)

APPLICATIONS OF STATISTICAL CONCEPTS IN THE BUSINESS WORLD

Many numerical data arrays in the business world closely resemble a uniform distribution and can be modeled as discussed in this chapter. Also, there are times when the normal approximation is useful for computing binomial probabilities. But the greatest application of the concepts in this chapter involves the many data collections found in the business world that approximate the shape of a normal curve.

When such a data array is found, percentages or probabilities can be easily determined using z values and the standard normal table. In fact, this method is so easy that it is frequently used even before the data distribution has been verified to resemble a normal curve. Obviously, it is important to verify the normality assumption if the results are to be accurate. Following are examples of variables that might follow a normal distribution and for which the techniques of this chapter are useful.

Accounting: accounts receivable, accounts payable, cash balance, value of inventory, bad debt expense.

Manufacturing: dollar value of inventory, percent defective in manufacturing process, number of days to receive shipment, number of overtime hours per week, hours of labor per unit, weights of manufactured parts, lengths of parts.

Finance: interest rate paid as deviation from prime rate, company's stock price as percentage of Dow Jones, ratio of annual dividends to earnings, several companies' price/earnings ratios.

Marketing: ages of credit card holders, numerical ratings of store performance by customers, monthly market share, percent of customers lost per month, miles traveled by customers to store.

Personnel: time with company, years of education, score on company entrance test, job performance rating, number of dependents, annual salary, sick days taken per year.

Quality control: time between defects, defective items per shift, percent out of tolerance, time until part failure, time until defective condition detected.

GLOSSARY

Continuous random variable A random variable measured on a numerical scale. An observation of the random variable can assume any value within some specified range.

Uniform distribution A distribution describing a random variable that is as likely to occur in one segment of a given size within a specified range as another.

Normal distribution A continuous distribution that has a bell shape and is determined by its mean and standard deviation.

Standard normal distribution A distribution that has a mean equal to 0 and a standard deviation equal to 1.

z value A measure of the number of standard deviations between the mean or center of a normal curve and the value of interest.

Chebyshev's theorem A rule stating that regardless of how the data values are distributed, at least $(1 - 1/k^2)$ of the values will lie within k standard deviations of the mean (for $k \ge 1$).

ared. What

on. If 110

ple of 300 probability

6% of the If 1,500 mine telene-person

ners have hat is the

re useful normal lications attributes can, and

iven the ons that r underal curve reaching

esented. inomial Poisson

KEY FORMULAS

Uniform probabilities

$$P(c \le X \le d) = \frac{d - c}{b - a} \tag{6.1}$$

Mean of uniform distribution

$$\mu = \frac{a+b}{2} \tag{6.2}$$

Standard deviation of uniform distribution

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} \tag{6.3}$$

Normal distribution probability density function

$$f(X) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(1/2) |\langle X - \mu \rangle / \sigma|^2}$$
(6.4)

z value or standardized value

$$z = \frac{X - \mu}{\sigma} \tag{6.5}$$

SOLVED EXERCISES

1. UNIFORM DISTRIBUTION

The Latah Creek Winery produces between 200 and 300 gallons of wine each day. The uniform distribution best describes this process.

- a. What is the average amount of wine produced each day?
- b. What is the amount of variability in the number of gallons of wine produced from day to day?
- c. In what percentage of days can production be expected to fall between 220 and 270
- d. What is the probability that production will be more than 280 gallons tomorrow?

a. Mean =
$$\frac{200 + 300}{2}$$
 = 250

b. Standard deviation =
$$\sqrt{\frac{(300 - 200)^2}{12}}$$
 = 28.9

c.
$$P(220 \le X \le 270) = \frac{270 - 220}{300 - 200} = \frac{50}{100} = .50$$

d. $P(280 \le X \le 300) = \frac{300 - 280}{300 - 200} = \frac{20}{100} = .20$

d.
$$P(280 \le X \le 300) = \frac{300 - 280}{200 - 200} = \frac{20}{100} = .20$$

2. NORMAL DISTRIBUTION

American Carpets has 2,000 accounts receivable. The mean and standard deviation are \$300 and \$50, respectively. Assume that the accounts are normally distributed.

- a. How many accounts exceed \$400?
- b. How many accounts are below \$250?

191

HAPTER SIX

..........

(6.1)

(6.2)

(6.3)

(6.4)

(6.5)

lay. The

d from

ind 270

rrow?

CONTINUOUS PROBABILITY DISTRIBUTIONS

- c. What is the probability that an account selected at random will be between \$200 and \$350?
- d. What percentage of the accounts are between \$325 and \$375?
- e. Forty percent of the accounts exceed what dollar amount? (Hint: Fifty percent of the accounts are for more than \$300.)
- f. Twenty percent of the accounts are below what dollar amount?

Solution:

a. z = (400 - 300)/50 = 2.0 → .4772 (Area between 300 and 400) (.5000 - .4772) = .0228 (Area above 400) 2,000 × .0228 = 45.6, or approximately 46 accounts

b. $z = (250 - 300)/50 = -1.0 \rightarrow .3413$ (Area between 250 and 300) (.5000 - .3413 = .1587 (Area below 250) 2,000 × .1587 = 317.4, or approximately 317 accounts

c. $z = (200 - 300)/50 = -2.0 \rightarrow .4772$ (Area between 200 and 300) $z = (350 - 300)/50 = 1.0 \rightarrow .3413$ (Area between 300 and 350) Sum = .8185

The probability is approximately .82.

d. $z = (375 - 300)/50 = 1.5 \rightarrow .4332$ (Area between 300 and 375) $z = (325 - 300)/50 = 0.5 \rightarrow .1915$ (Area between 300 and 325) Difference = .2417 (Area between 325 and 375)

The percentage is approximately 24%.

e. $(.5000 - .4000) = .1000 \rightarrow z = 0.25$ $X = \mu + z\sigma = 300 + 0.25(50) = 312.5$

Forty percent of the accounts are worth more than \$312.50.

f. $(.5000 - .2000) = .3000 \rightarrow z = -0.84$ $X = \mu + z\sigma = 300 + (-0.84)(50) = 258$

Twenty percent of the accounts are worth less than \$258

3. NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

Colorado Power Company provides lower rates to customers who agree to use energy mainly at off-peak hours. Thirty percent of their customers take advantage of these savings. The consumer affairs department has conducted a focus group and is preparing to conduct a random telephone survey of 500 customers. The department supervisor wants to make sure that the group will contain a sufficient proportion of off-peak users.

- a. What is the probability of obtaining fewer than 150 off-peak users in the telephone survey?
- b. What is the smallest number of off-peak users likely to be included in this sample? (Hint: Use 3 standard deviations below the mean, which will include .9987 of the area under the curve.)

Solution:

a. $\mu = np = 500(.30) = 150$

Since 50% of a normal curve lies below the mean, the probability of obtaining fewer than 150 customers is 50%.

b. $\sigma = \sqrt{np(1-p)} = \sqrt{500(.30)(.70)} = 10.25$ $x = \mu + z\sigma = 150 + (-3.0)(10.25) = 119.25$

The probability of fewer than 119 off-peak users in the sample is .0013.

EXERCISES

- 37. Draw a normal curve and shade in the area under the curve for each of the following:
 - a. The area between the mean and z = 2.50
 - b. The area between the mean and z = -0.75

ion are

ion are

- The area between z = -1.55 and z = 0.89
- d. The area between z = -0.38 and z = -2.18
- e. The area above z = 2.68
- f. The area below z = -1.15
- Find the probability that an item would fall in each of the shaded intervals drawn in Exercise 37.
- 39. Find the z value that corresponds to each of the areas described:
 - a. Fifty percent of the items lie above this z value.
 - b. Thirty percent of the items lie below this z value.
 - c. The chances are 1 in 10 that an item selected at random will deviate from the mean by more than this z value.
 - d. Five percent of the items will lie more than this distance (measured in terms of a z value) from the mean.
- The total time needed to process a mortgage loan application at Baneshares Mortgage Company is uniformly distributed between five and nine days.
 - a. What is the average amount of loan processing time?
 - b. What is the amount of variability in the loan processing time?
 - e. What percentage of loans are processed within eight days?
 - d. What is the probability that a loan will take more than six days to be processed?
- 41. Mary Walsh is studying the records of the Valley Appliance Company to determine the adequacy of the present warranty policy for refrigerators. She finds that the length of time Valley Appliance refrigerators operate before requiring repairs is normally distributed with mean 5.7 years and standard deviation 2.1 years. The company presently repairs free any machine that fails to work properly within a two-year period after purchase. What percentage of Valley Appliance refrigerators require these free repairs?
- 42. The Barcellos' Tum Tum Resort staff keeps track of the number of days each guest stays. They have discovered that this variable is normally distributed with a mean of 9 and a standard deviation of 3. The forecast for next month indicates that 500 guests are expected.
 - a. How many guests are expected to stay fewer than 5 days?
 - b. How many guests are expected to stay more than 15 days?
 - c. How many guests are expected to stay between 6 and 12 days?
 - d. Fifteen percent of the guests leave after how many days?
 - e. Seventy-five percent of the guests stay longer than how many days?
- 43. The label on boxes of Spic and Clean, a laundry detergent, indicates that the net weight is 27 ounces. A machine fills these boxes in a uniformly distributed manner, with the smallest box containing 26 ounces and the largest 28 ounces. Quality control accepts boxes filled within 0.9 ounces of the amount stated on the box.
 - a. How much variability is there in the amount put in the boxes?
 - b. What is the probability that a box will be filled with between 26.8 and 27.5 ounces?
 - e. What is the probability that a box will fail to meet the quality control standard?
- 44. In a large L-Mart discount store, the maintenance department has been instructed to replace all of the light bulbs at the same time. Past experience indicates that the life of light bulbs is normally distributed with a mean life of 750 hours and a standard deviation of 40 hours. When should the lights be replaced so that no more than 7% burn out?
- 45. The accountant for the Forrest Paint Company, Don Horner, has a history of making errors on 6% of the invoices he processes. Forrest Paint processed 400 invoices last month.
 - a. How many invoices would you expect to have errors?
 - b. What is the probability that Don made fewer than 20 errors?
 - c. What is the probability that Don made more than 30 errors?
- 46. Jake runs the concession stand for a minor league baseball team in the Chicago area. He knows that the sales of hot dogs are normally distributed with an average of 600 hot dogs per day and a standard deviation of 45. Jake begins to consider how much to stock in his

business after reading an article in Sport magazine (March 1989) about baseball in the Chicago area. This article predicted good crowds for Chicago teams, and Jake thinks this will affect his own business.

Jake wants to make sure that he doesn't run out of hot dogs. How many hot dogs should he order so that the probability of running out is less than 1% each day?

- The rate at which a swimming pool is filled is between 32 and 39 gallons per minute. (Assume a uniform distribution.)
 - a. What is the average rate at which the swimming pool is filled?
 - b. What is the probability that the rate at which the pool is filled at any one time is below 35 gallons per minute?
- 48. On the basis of a study, automobile inspectors in New Mexico have determined that 11% of all cars coming in for annual inspection fail. What is the probability that more than 20 cars out of the next 130 cars tested will fail the inspection?
- 49. An automatic machine produces bearings, 2% of which are defective. If a sample of 600 bearings is drawn at random from the production of this machine, what is the probability of observing 6 or fewer defectives?
- 50. The personnel director of Kershaw's, Inc., Kelly Moreland, needs to administer an aptitude test to 30 job applicants. Kelly has just read an article about labor shortages and how managers cope with this problem in *Personnel Management* (May 1989 issue). This article made Kelly think about how to improve the selection process for hirees, and she has bought an aptitude exam from a national company for use at Kershaw's.

The booklet for the test informs her that the average length of time required to complete the exam is 120 minutes with a standard deviation of 25 minutes. When should she terminate the examination if she wishes to allow sufficient time for 80% of the applicants to finish? (Assume that the test times are normally distributed.)

- 51. Associated Investment Advisors sells annuities based on annual payout during the lifetimes of the participants in the plan. They have determined that the lengths of life of participants in the plan are approximately normally distributed with a mean of 72 years and a standard deviation of 4.7 years.
 - a. What is the probability that a plan participant will receive payments beyond age 65? Beyond age 70? Beyond age 75? Beyond age 80?
 - b. Sixty percent of the plan participants are beyond what age?
- 52. The city planner, Terry Novak, knows that mean family income is \$28,000 with a standard deviation of \$9,000 in Bonners Ferry, Kentucky, a community of 20,000 households. He also knows that family income is not a normally distributed variable. Terry is applying for a grant and must answer the following question: At least what percent of incomes lie within \$10,000 of the mean income?
- 53. Innovative Technology Systems manufactures electronic cables for computers. Innovative has found from past experience that 97% of the cables shipped are acceptable. They ship most of their cables to Compaq Computers, and The Wall Street Journal (November 3, 1989) has just indicated that Compaq's stock plunged after it reported lower-than-expected fourth-quarter profits. This has Innovative concerned about the quality of their cables.

A shipment of 400 cables has just been sent to Compaq. Innovative's contract calls for at least 380 acceptable cables. What is the probability that this shipment will fulfill the contract conditions?

54. You are finishing your business degree at State University. You have worked part-time at Ranch Thrift, a large supermarket, for the past three years. Your boss, Archie Copeland, knows that you are completing your degree and asks you to study the average customer expenditure per visit. You sample 500 customer expenditures and determine that this variable is normally distributed with an average expenditure of \$31.28 and a standard deviation of \$7.29. Write your boss a memo analyzing the average customer expenditure per visit.

vals drawn in

om the mean

terms of a z

res Mortgage

cessed?

etermine the ength of time tributed with mairs free any at percentage

guest stays. of 9 and a re expected.

e net weight et, with the eccepts boxes

1.5 ounces? idard?

structed to t the life of d deviation m out?

of making ast month.

o area. He 0 hot dogs tock in his

EXTENDED EXERCISES

55. IRS REFUNDS

An IRS regional office has plotted a large sample of tax refund amounts and found that they form a bell-shaped curve that is symmetrical about a center line. Most refund values are fairly close to this center value, suggesting a normal distribution. The average refund amount for the office is calculated to be \$748 with a standard deviation of \$124. Since the data plot is so close to a normal curve, it is decided that probability or percentage questions can be answered using the standard normal table.

The regional office is specifically interested in the proportion of large refund amounts. A refund greater than \$1,000 is considered large, and the office would like to know what percentage of current refunds, based on the normal distribution, are in excess of this amount.

A second question involves the office's concern with new withholding guidelines for taxpayers. It has been estimated that the average refund amount will rise by \$150 after these guidelines go into effect, and the IRS office wonders how the percentage of large refunds will be affected.

The office decides to assume a normal distribution, with a mean of \$748 and a standard deviation of \$124, to find the percentage of refunds over \$1,000 that might be expected based on current guidelines. The z value is calculated, and the resulting area is subtracted from .50 to determine the area in the standard normal upper tail:

$$z = (1,000 - 748)/124 = 2.03 \rightarrow .4788$$
 (Area between 748 and 1,000)
(.5000 - .4788) = .0212 (Area above 1,000)

About 2% of the refunds under the current withholding guidelines can be expected to exceed \$1,000.

With the new guidelines in effect, a \$150 increase in the average refund amount can be expected. This will increase the average refund to (748 + 150) = 898. It is assumed that the standard deviation of refund balances will remain the same, \$124, even though the mean has shifted. This assumption is based on the notion that even though the average increases, the degree of variability in refund amounts will stay the same. Under this assumption, the percentage of the normal curve above \$1,000 can be found using the z value and the standard normal table:

$$z = (1,000 - 898)/124 = 0.82 \rightarrow .2939$$
 (Area between 898 and 1,000)
(.5000 - .2939) = .2061 (Area above 1,000)

The increase in average refund amount raises the percentage of refunds over \$1,000 to about 21%. This increase from 2% to 21% is of great importance to the IRS office in planning for funding requirements and clerical time. The information has resulted from the assumption that refunds can be modeled with the normal distribution, giving the office important lead time in dealing with the withholding policy changes.

 What would happen to the percent of refunds over \$1,000 if the average refund amount actually dropped \$75?

2. After the 2% figure has been computed under the current withholding policy, how could the assumption of a normal distribution be checked?

56. RANCH LIFE INSURANCE COMPANY OF FLORIDA

Kim Carter, statistician for the Ranch Life Insurance Company of Florida, notices that monthly premium receipts appear to follow a normal distribution. After plotting the receipts for a typical month and observing their distribution, Kim decides that the normality assumption is reasonable.

Ranch is interested in determining the percentage of monthly receipts that fall into the range \$125 to \$175, since such amounts are difficult to handle. They are not large enough

found that and values age refund 24. Since

amounts. now what ess of this

percentage

slelines for \$150 after se of large

a standard e expected subtracted

spected to

nount can s assumed on though he average Inder this sing the z

\$1,000 to office in dted from the office

d amount

licy, how

otices that ne receipts normality

ll into the

to warrant investment attention, unless there are many of them, nor are they small enough to ignore by putting them into the cash account.

To find the percentage of accounts in this range, Kim must compute the mean and standard deviation of monthly premium receipts. She determines these values to be \$100 and \$38, respectively.

- What percent of monthly receipts can be expected to fall into the range \$125 to \$175?
- 2. Kim predicts that within two years the average monthly premium will rise to \$150. Assuming the standard deviation remains the same, how would this change the answer to Question 1?
- 3. Describe how the assumption of normality might be verified in this case.

57. SPECIFIC ELECTRIC

Jess Franklin of the finance department for the Specific Electric Appliance Manufacturing Company has been studying the interest rates charged on its short-term loans by the company's banks. For comparison purposes, he has been following the T-bill rate for several months. The most recent yield, as reported in *The Wall Street Journal* (February 23, 1990), is 7.80%. He has also been carefully watching the rate charged by Specific's bank.

Jess has made the assumption that his bank rate, which varies from day to day, can be described by a normal curve and has used normal curve procedures to compute various probabilities. These probabilities are used in preparing for market rate shifts under an acceptable level of risk.

In making a presentation to top management of the firm, Jess briefly mentions the normality assumption and then shows the computations for market rate probabilities and the preparations based on these probabilities. During the presentation, the company president, Mary Montana, interrupts with the following statement: "I don't know about your assumption of a normal distribution for daily interest rates. It seems to me the assumption is not valid, since rates often seem to be on a long-term increase or decrease. If that's true, the assumption of a normal distribution may be valid, but not the assumption of an unchanging average."

- 1. What are the strengths in Mary's statement?
- 2. What are the weaknesses in Mary's statement?
- 3. Is there any way Mary's concern about a shifting mean could be incorporated into Jess's calculations?

58. GMAT Scores For City University

Professors Williams, Ross, Stanley, and Richards are serving on the MBA Committee at City University. The current task of the committee is to set a cutoff level for admitting students into the MBA program using the GMAT (Graduate Management Admissions Test). Currently, the mean test score nationally is 475 with a standard deviation of 75. The scores are thought to be approximately normally distributed.

The GMAT has been administered to the last 100 applicants for City's MBA program. Professor Ross suggests that these 100 scores should be analyzed to determine if applicants to the MBA program at City University are similar to those nationally. If the sample is similar, Professor Ross proposes that a score be chosen that would permit 25% of the applicants to be admitted to the program. The 100 GMAT scores are:

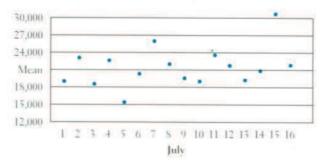
568 606 365 422 505 600 461 412	395 486 415 506 429 568 564 483	489 514 575 542 564 450 519 452	288 497 432 452 438 580 375 389	396 399 378 579 350 450 553 611	442 562 403 400 595 423 543 569	672 507 518 429 476 598 480 368	492 457 476 383 469 458 322 469	465 449 355 569 462 494 490 348	523 553 470 381 446 582 436 522	494 492 512 514 565 623 446 479	513 330 527 465 450 401 465 473	486 375 366 439
--	--	--	--	--	--	--	--	--	--	--	--	--------------------------

- CHAPTER SIX
- 1. Are City University applicants similar to the national average?
- 2. If Professor Ross's proposal is accepted, what cutoff score should be used for admittance
- 3. Professor Richards feels that one third of the applicants should be admitted. What cutoff score would he propose?

59. EUREKA DAIRY

Jim Black is in charge of quality control at the Eureka Dairy, located in Deer Park Georgia. Jim is very concerned about the bacteria count per cubic centimeter of grade A milk. On July 15 the bacteria count was 31,000. Jim's problem is to determine whether something is wrong with the production process. The Eureka Dairy over a long period of time has regulated its processing to achieve a mean bacteria count of 21,000 with a standard deviation of 3,000. A distribution plotted from the bacteria counts over a long period of time is normal in form. Each day the count is plotted on a chart similar to that shown in Figure 6.18. Jim needs to determine whether this chart is useful.

FIGURE 6.18 Bacteria Counts in Grade A Milk



- 1. What are the chances in 1,000 that the bacteria count will exceed 30,000 if the day to-day variation shown in Figure 6.18 is purely random?
- 2. What is the probability of a bacteria count of 31,000 or more?
- 3. Which is the more reasonable assumption: that a bacteria count of 31,000 might reasonably be expected, or that something is wrong? Explain your answer.
- Is the charting device shown in Figure 6.18 helpful in maintaining a high standard of quality? Explain your answer.
- 5. Why would it be important for this dairy to find the mean value of these bacteria counts every month?

MICRO COMPUTER PACKAGE

The micro package Computerized Business Statistics can be used to solve normal curve problems In Situation 6.2, Applied Business Systems, Inc., needed to know what percent of its order exceeded 135 miles. Applied's analyst knows that shipping distances are normally distributed with a mean of 125 miles and a standard deviation of 28 miles.

Computer Solution:

From the main menu of Computerized Business Statistics, a 5 is selected, which indicates Probability Distributions. The probability distributions menu is shown on the screen. This menu was presented at the end of Chapter 5.

This problem involves the normal distribution. Since Applied needed to know what percent of its orders exceeded 135 miles, the arrow is moved to normal $P(X \ge a)$. Instructions now appear on the screen, and data are entered in response to each instruction as follows:

HAPTER SIX

dmittance? tted. What

Deer Park, of grade A ne whether CONTINUOUS PROBABILITY DISTRIBUTIONS

197

Enter Mean and Press ← 125
Enter Standard Deviation and Press ← 28
Enter A and Press ← 135

The normal probability distribution for $\mu = 125$ and $\sigma = 28$ is shown on the screen, and the correct answer for $P(X \ge 135)$ is .3601.

g period of a standard g period of that shown

if the day-

000 might

h standard

se bacteria

problems. f its orders distributed

indicates
This menu

at percent tions now

\$: