7 Normal Distribution

Continuous random variables and their associated density functions arise whenever our experimental data are defined over a continuous sample space. Therefore, whenever we measure time intervals, weights, heights, volumes, and so forth, our underlying population is described by a continuous distribution. Just as there are several special discrete probability distributions there are also numerous types of continuous distributions whose graphs may display varying amounts of skewness or in some cases may be perfectly symmetric. Among these, by far the most important is the continuous distribution whose graph is a symmetric bell-shaped curve extending indefinitely in both directions. It is this distribution that provided a basis upon which much of the theory of statistical inference has been developed.

.1

ormal Curve

The most important continuous probability distribution in the entire field of statistics is the **normal distribution.** Its graph, called the **normal curve,** is the bell-shaped curve of Figure 7.1 that describes so many sets of data that occur in nature, industry, and research. In 1733, DeMoivre derived the mathematical equation of the normal curve. The normal distribution is often referred to as the **Gaussian distribution** in honor of Gauss (1777–1855), who also derived its equation from a study of errors in repeated measurements of the same quantity.

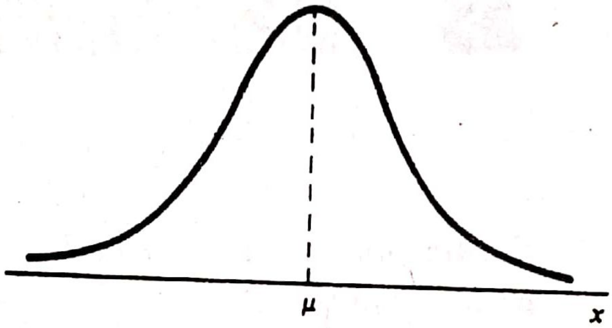

FIGURE 7.1
The normal curve.

A continuous random variable X having the bell-shaped distribution of Figure 7.1 is called a **normal random variable.** The mathematical equation for the probability distribution of the normal variable depends upon the two parameters μ and σ, its mean and standard deviation. Hence we denote the values of the density function of X by $n(x; \mu, \sigma)$.

DEFINITION

Normal Curve. If X is a normal random variable with mean μ and variance σ^2, then the equation of the *normal curve* is

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad \text{for } -\infty < x < \infty,$$

where $\pi = 3.14159 \ldots$ and $e = 2.71828 \ldots$.

Once μ and σ are specified, the normal curve is completely determined. For example, if $\mu = 50$ and $\sigma = 5$, then the ordinates of $n(x; 50, 5)$ can easily be computed for various values of x and the curve drawn. In Figure 7.2 we have sketched two normal curves having the same standard deviation but

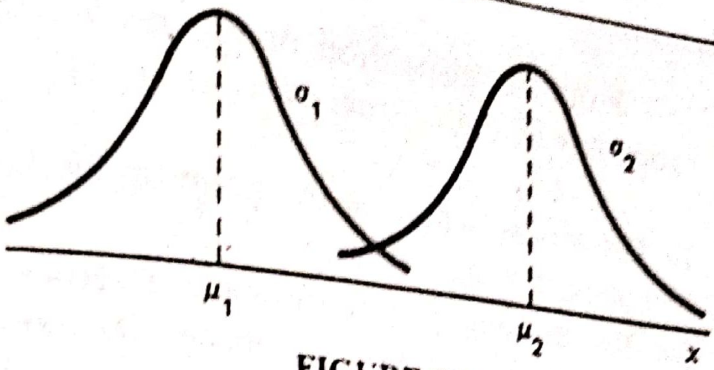

FIGURE 7.2
Normal curves with $\mu_1 \neq \mu_2$ and $\sigma_1 = \sigma_2$.

different means. The two curves are identical in form but are centered at different positions along the horizontal axis.

In Figure 7.3 we have sketched two normal curves with the same mean but different standard deviations. This time we see that the two curves are centered at exactly the same position on the horizontal axis, but the curve with the larger standard deviation is lower and spreads out farther. Remember that the area under a probability curve must be equal to 1, and therefore the more variable the set of observations, the lower and wider the corresponding curve will be.

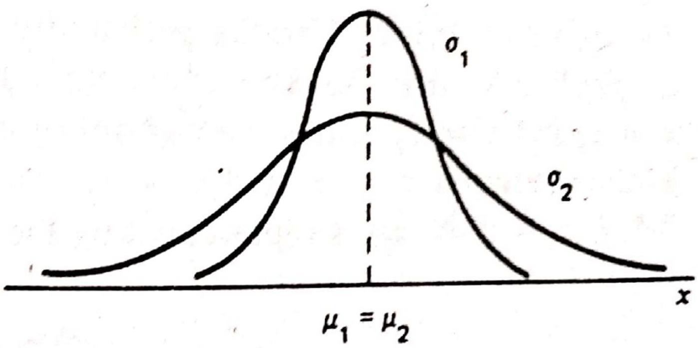

FIGURE 7.3
Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

Figure 7.4 shows the results of sketching two normal curves having different means and different standard deviations. Clearly, they are centered at different positions on the horizontal axis and their shapes reflect the two different values of σ.

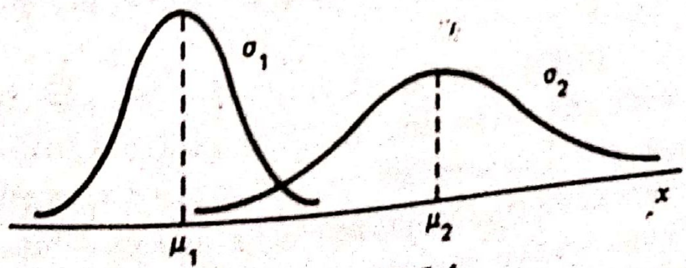

FIGURE 7.4
Normal curves with $\mu_1 \neq \mu_2$ and $\sigma_1 < \sigma_2$.

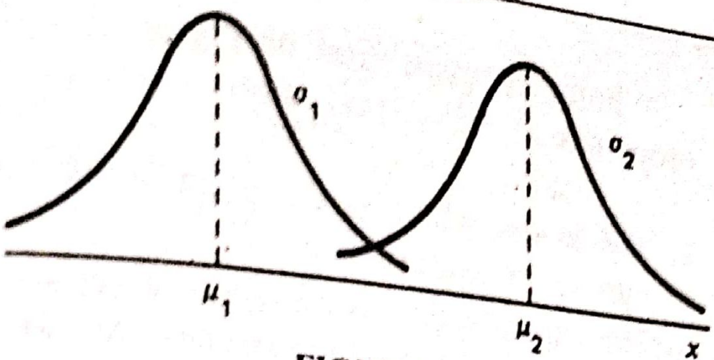

FIGURE 7.2
Normal curves with $\mu_1 \neq \mu_2$ and $\sigma_1 = \sigma_2$.

different means. The two curves are identical in form but are centered at different positions along the horizontal axis.

In Figure 7.3 we have sketched two normal curves with the same mean but different standard deviations. This time we see that the two curves are centered at exactly the same position on the horizontal axis, but the curve with the larger standard deviation is lower and spreads out farther. Remember that the area under a probability curve must be equal to 1, and therefore the more variable the set of observations, the lower and wider the corresponding curve will be.

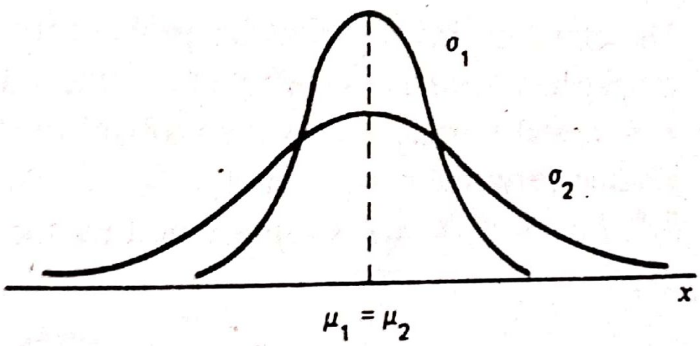

FIGURE 7.3
Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

Figure 7.4 shows the results of sketching two normal curves having different means and different standard deviations. Clearly, they are centered at different positions on the horizontal axis and their shapes reflect the two different values of σ.

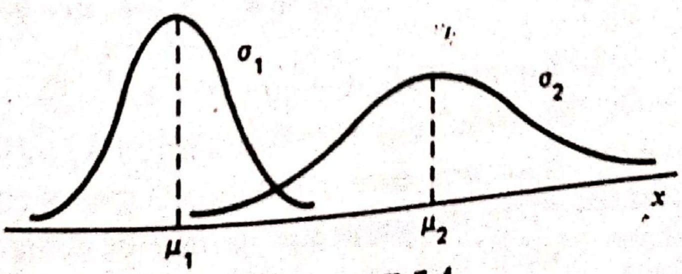

FIGURE 7.4
Normal curves with $\mu_1 \neq \mu_2$ and $\sigma_1 < \sigma_2$.

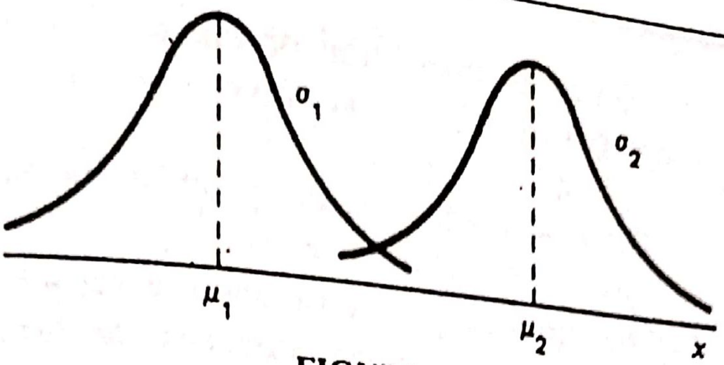

FIGURE 7.2
Normal curves with $\mu_1 \neq \mu_2$ and $\sigma_1 = \sigma_2$.

different means. The two curves are identical in form but are centered at different positions along the horizontal axis.

In Figure 7.3 we have sketched two normal curves with the same mean but different standard deviations. This time we see that the two curves are centered at exactly the same position on the horizontal axis, but the curve with the larger standard deviation is lower and spreads out farther. Remember that the area under a probability curve must be equal to 1, and therefore the more variable the set of observations, the lower and wider the corresponding curve will be.

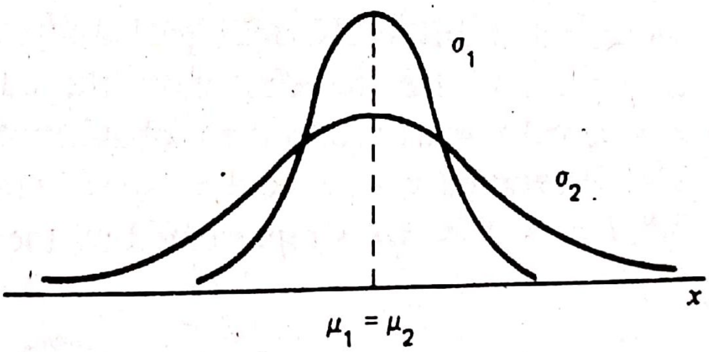

FIGURE 7.3
Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$.

Figure 7.4 shows the results of sketching two normal curves having different means and different standard deviations. Clearly, they are centered at different positions on the horizontal axis and their shapes reflect the two different values of σ.

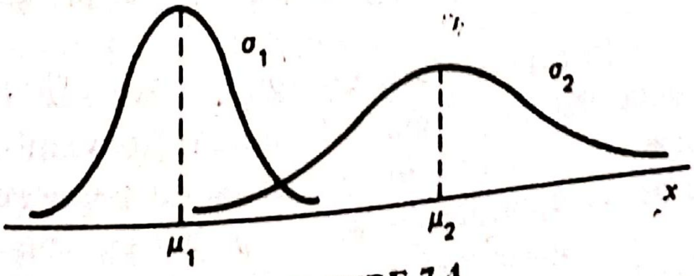

FIGURE 7.4
...with $\mu_1 \neq \mu_2$ and $\sigma_1 < \sigma_2$.

From an inspection of Figures 7.1 through 7.4, we list the following properties of the normal curve:

1. The *mode*, which is the point on the horizontal axis where the curve is a maximum, occurs at $x = \mu$.
2. The curve is symmetric about a vertical axis through the mean μ.
3. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
4. The total area under the curve and above the horizontal axis is equal to 1.

Many random variables have probability distributions that can be described adequately by the normal curve once μ and σ^2 are specified. In this chapter we assume that these two parameters are known, perhaps from previous investigations. Later, in Chapter 8, we consider methods of estimating μ and σ^2 from the available experimental data.

7.2 Areas Under the Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates $x = x_1$ and $x = x_2$ equals the probability that the random variable X assumes a value between $x = x_1$ and $x = x_2$. Thus, for the normal curve in Figure 7.5, $P(x_1 < X < x_2)$ is represented by the area of the shaded region.

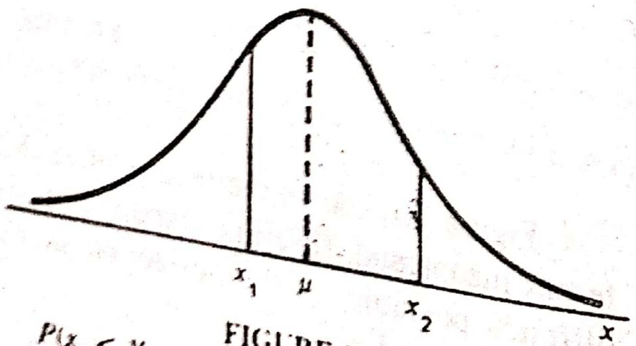

FIGURE 7.5
$P(x_1 < X < x_2)$ = area of the shaded region.

In Figures 7.2 through 7.4 we saw how the normal curve is dependent upon the mean and the standard deviation of the distribution under investigation. The area under the curve between any two ordinates must then also depend upon the values of μ and σ. This is evident in Figure 7.6, where we have shaded regions corresponding to $P(x_1 < X < x_2)$ for two curves with different means and variances. The $P(x_1 < X < x_2)$, where X is the random

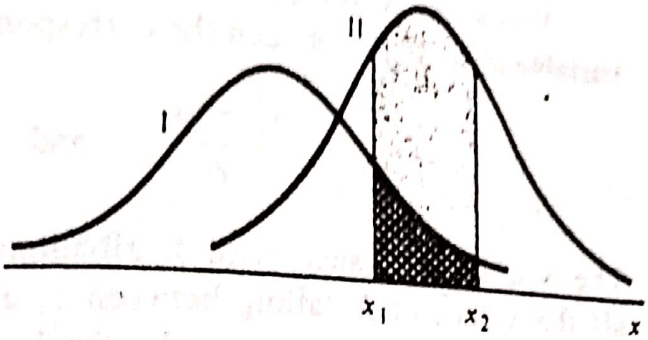

FIGURE 7.6
$P(x_1 < X < x_2)$ for different normal curves.

variable describing distribution I, is indicated by the crosshatched area. If X is the random variable describing distribution II, then the $P(x_1 < X < x_2)$ is given by the entire shaded region. Obviously, the two shaded regions are different in size; therefore, the probability associated with each distribution will be different.

It would be a hopeless task to attempt to set up separate tables of normal curve areas for every conceivable value of μ and σ. Yet we must use tables if we hope to avoid the use of integral calculus. Fortunately, we are able to transform all the observations of any normal random variable X to a new set of observations of a normal random variable Z with mean zero and variance 1. This can be done by means of the transformation

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

The mean of Z is zero, since

$$E(Z) = \frac{1}{\sigma} E(X - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0,$$

and the variance is 1,

$$\sigma_Z^2 = \sigma_{(X-\mu)/\sigma}^2 = \sigma_{X/\sigma}^2 = \frac{1}{\sigma^2}\sigma_X^2 = \frac{\sigma^2}{\sigma^2} = 1.$$

ITION

Standard Normal Distribution. The distribution of a normal random variable with mean zero and standard deviation equal to 1 is called a *standard normal distribution*.

Scanned with CamScanner

Whenever X is between the values $x = x_1$ and $x = x_2$, the random variable Z will fall between the corresponding values

$$z_1 = \frac{x_1 - \mu}{\sigma} \quad \text{and} \quad z_2 = \frac{x_2 - \mu}{\sigma}.$$

The original and transformed distributions are illustrated in Figure 7.7. Since all the values of X falling between x_1 and x_2 have corresponding z values between z_1 and z_2, the area under the X curve between the ordinates $x = x_1$ and $x = x_2$ in Figure 7.7 equals the area under the Z curve between the transformed ordinates $z = z_1$ and $z = z_2$. Hence we have

$$\boxed{P(x_1 < X < x_2) = P(z_1 < Z < z_2).}$$

We have now reduced the required number of tables of normal-curve areas to one—that of the standard normal distribution. Table A.4 gives the area under the standard normal curve corresponding to $P(Z < z)$ for values of z from -3.49 to 3.49. To illustrate the use of this table, let us find the probability that Z is less than 1.74. First we locate a value of z equal to 1.7 in the left column and then move across the row to the column under 0.04, where we read 0.9591. Therefore, $P(Z < 1.74) = 0.9591$.

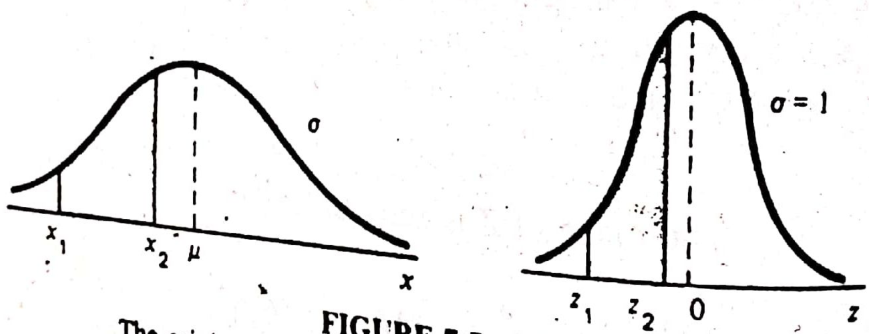

FIGURE 7.7
The original and transformed normal populations.

?

Occasionally, we are required to find a value of z corresponding to a specified probability that falls between values listed in Table A.4 (see Example 3). For convenience, we shall always choose the z value corresponding to the tabular value that comes closest to the specified probability. However, if the given probability falls midway between two tabular values, we shall choose for z the value falling midway between the corresponding values of z. For instance, to find the z value corresponding to a probability of 0.7975, which falls between 0.7967 and 0.7995 in Table A.4, we choose $z = 0.83$, since 0.7975 is closer to 0.7967. On the other hand, for a probability of 0.7981, which falls midway between 0.7967 and 0.7995, we take ...

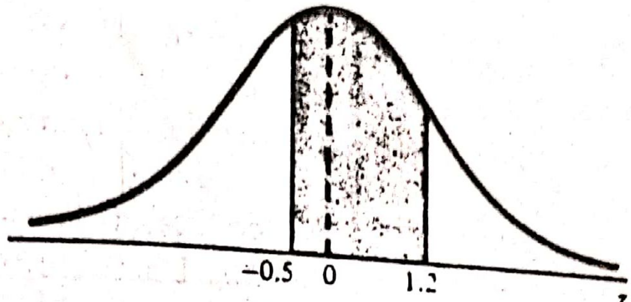

FIGURE 7.8
Area for Example 1.

Example 1. Given a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

Solution. The z values corresponding to $x_1 = 45$ and $x_2 = 62$ are

$$z_1 = \frac{45 - 50}{10} = -0.5,$$

$$z_2 = \frac{62 - 50}{10} = 1.2.$$

Therefore,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2).$$

The $P(-0.5 < Z < 1.2)$ is given by the area of the shaded region in Figure 7.8. This area may be found by subtracting the area to the left of the ordinate $z = -0.5$ from the entire area to the left of $z = 1.2$. Using Table A.4, we have

$$\begin{aligned} P(45 < X < 62) &= P(-0.5 < Z < 1.2) \\ &= P(Z < 1.2) - P(Z < -0.5) \\ &= 0.8849 - 0.3085 \\ &= 0.5764. \end{aligned}$$

Example 2. Given a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

Solution. The normal probability distribution showing the desired area is given in Figure 7.9. To find the $P(X > 365)$, we need to evaluate the area under the normal curve to the right of $x = 362$. This can be done by

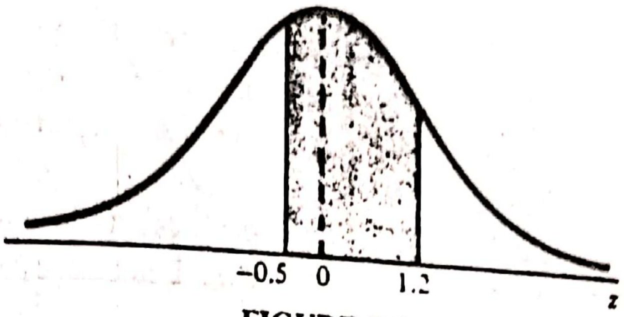

FIGURE 7.8
Area for Example 1.

Example 1. Given a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

Solution. The z values corresponding to $x_1 = 45$ and $x_2 = 62$ are

$$z_1 = \frac{45 - 50}{10} = -0.5,$$

$$z_2 = \frac{62 - 50}{10} = 1.2.$$

Therefore,

$$P(45 < X < 62) = P(-0.5 < Z < 1.2).$$

The $P(-0.5 < Z < 1.2)$ is given by the area of the shaded region in Figure 7.8. This area may be found by subtracting the area to the left of the ordinate $z = -0.5$ from the entire area to the left of $z = 1.2$. Using Table A.4, we have

$$\begin{aligned} P(45 < X < 62) &= P(-0.5 < Z < 1.2) \\ &= P(Z < 1.2) - P(Z < -0.5) \\ &= 0.8849 - 0.3085 \\ &= 0.5764. \end{aligned}$$

Example 2. Given a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

Solution. The normal probability distribution showing the desired area is given in Figure 7.9. To find the $P(X > 365)$, we need to evaluate the area under the normal curve to the right of $x = 362$. This can be done by

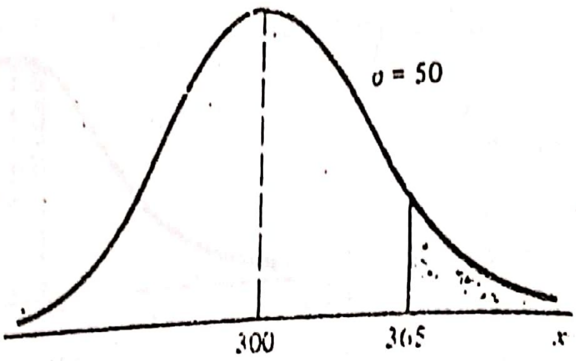

FIGURE 7.9
Area for Example 2.

transforming $x = 362$ to the corresponding z value, obtaining the area to the left of z from Table A.4, and then subtracting this area from 1. We find that

$$z = \frac{362 - 300}{50} = 1.24.$$

Hence

$$\begin{aligned} P(X > 362) &= P(Z > 1.24) \\ &= 1 - P(Z < 1.24) \\ &= 1 - 0.8925 \\ &= 0.1075. \end{aligned}$$

Example 3. Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value x that has (a) 38% of the area below it and (b) 5% of the area above it.

Solution. The preceding two examples were solved by going first from a value of x to a z value and then computing the desired area. In this example we reverse the process and begin with a known area or probability, find the z value, and then determine x by rearranging the formula

$$z = \frac{x - \mu}{\sigma} \quad \text{to give} \quad x = \sigma z + \mu.$$

(a) An area of 0.38 to the left of the desired x value is shown shaded in Figure 7.10. We require a z value that leaves an area of 0.38 to the left. From Table A.4 we find $P(Z < -0.31) = 0.38$ so that the desired

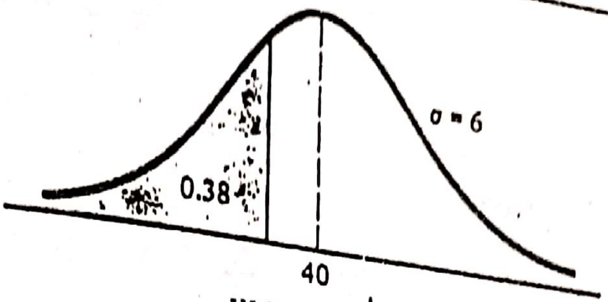

FIGURE 7.10
Area for Example 3(a).

z value is -0.31. Hence

$$x = (6)(-0.31) + 40$$
$$= 38.14.$$

(b) In Figure 7.11 we shade an area equal to 0.05 to the right of the desired x value. This time we require a z value that leaves an area of 0.05 to the right and hence an area of 0.95 to the left. Again from Table A.4 we find $P(Z < 1.645) = 0.95$ so that the desired z value is 1.645 and

$$x = (6)(1.645) + 40$$
$$= 49.87.$$

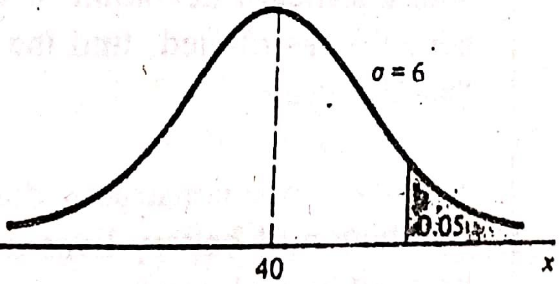

FIGURE 7.11
Area for Example 3(b).

The percentages associated with the Empirical Rule as stated in Section 3.3 were determined theoretically by using the normal-curve areas of Table A.4. For example, the probability of a normal random variable X with mean μ and variance σ^2 assuming a value between $x_1 = \mu - 2\sigma$ and $x_2 = \mu + 2\sigma$ is equal to the area under a standard normal curve between the ordinates $z = z_1$ and $z = z_2$, where

$$z_1 = \frac{(\mu - 2\sigma) - \mu}{\sigma} = -2$$

and

$$z_2 = \frac{(\mu + 2\sigma) - \mu}{\sigma} = 2.$$

Hence

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(-2 < Z < 2) \\ &= P(Z < 2) - P(Z < -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544, \end{aligned}$$

which is equivalent to stating that 95.44% of the measurements of a norm random variable fall within the interval $\mu \pm 2\sigma$.

7.3 Applications of the Normal Distribution

Some of the many problems in which the normal distribution is applicable treated in the following examples. The use of the normal curve to approxim binomial probabilities will be considered in Section 7.4.

Example 4. A certain type of storage battery lasts on the average 3.0 yea with a standard deviation of 0.5 year. Assuming that the battery lives normally distributed, find the probability that a given battery will last l than 2.3 years.

Solution. First construct a diagram such as Figure 7.12, showing the gi distribution of battery lives and the desired area. To find the $P(X < 2$ we need to evaluate the area under the normal curve to the left of 2.3. T is accomplished by finding the area to the left of the corresponding z val Hence we find that

$$z = \frac{2.3 - 3}{0.5} = -1.4,$$

and then, using Table A.4, we have

$$\begin{aligned} P(X < 2.3) &= P(Z < -1.4) \\ &= 0.0808 \end{aligned}$$

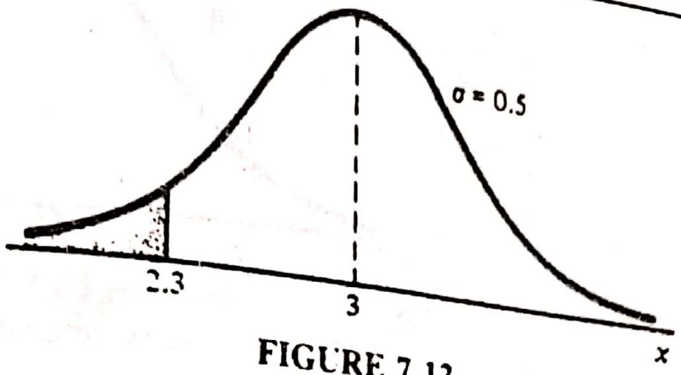

FIGURE 7.12
Area for Example 4.

Example 5. An electrical firm manufactures light bulbs that have a length of life that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

Solution. The distribution of light bulbs is illustrated in Figure 7.13. The z values corresponding to $x_1 = 778$ and $x_2 = 834$ are

$$z_1 = \frac{778 - 800}{40} = -0.55,$$

$$z_2 = \frac{834 - 800}{40} = 0.85.$$

Hence

$$\begin{aligned} P(778 < X < 834) &= P(-0.55 < Z < 0.85) \\ &= P(Z < 0.85) - P(Z < -0.55) \\ &= 0.8023 - 0.2912 \\ &= 0.5111. \end{aligned}$$

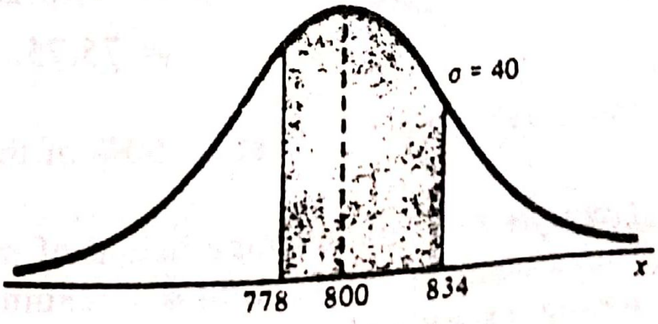

FIGURE 7.13
Area for Example 5.

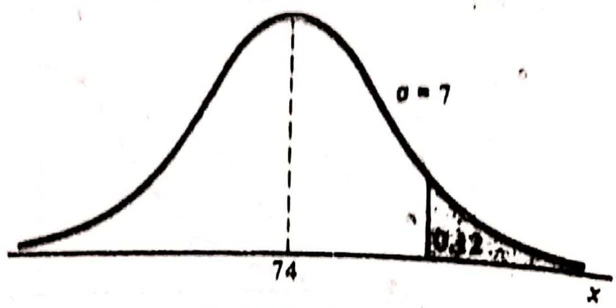

FIGURE 7.14
Area for Example 6.

Example 6. On an examination the average grade was 74 and the standard deviation was 7. If 12% of the class are given A's, and the grades are curved to follow a normal distribution, what is the lowest possible A and the highest possible B?

Solution. In this example we begin with a known area or probability, find the z value, and then determine x from the formula $x = \sigma z + \mu$. An area of 0.12, corresponding to the fraction of students receiving A's, is shaded in Figure 7.14. We require a z value that leaves 0.12 of the area to the right and hence an area of 0.88 to the left. From Table A.4, $P(Z < 1.175) = 0.88$ so that the desired z value is 1.175. Hence

$$\begin{aligned} x &= (7)(1.175) + 74 \\ &= 82.225. \end{aligned}$$

Therefore, the lowest A is 83 and the highest B is 82.

Example 7. Refer to Example 6 and find D_6.

Solution. The sixth decile, D_6, is the x value below which [illegible] of the area lies as shown by the shaded region in Figure 7.15. From Table [illegible] .4 we find $P(Z < 0.25) = 0.6$ so that the desired z value is 0.25. Now

$$\begin{aligned} x &= (7)(0.25) + 74 \\ &= 75.75. \end{aligned}$$

Hence $D_6 = 75.75$. That is, 60% of the grades are 75 or less.

Example 8. If the average height of miniature poodles is 30 centimeters, with a standard deviation of 4.1 centimeters, what percentage of minia[ture] poodles exceeds 35 centimeters in height, assuming that the heights foll[ow]

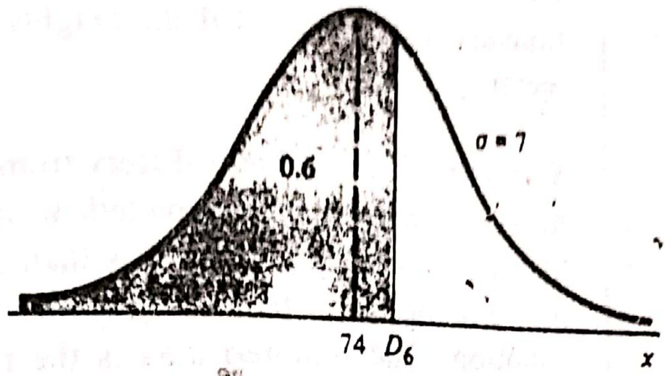

FIGURE 7.15
Area for Example 7.

a normal distribution and can be measured to any desired degree of accuracy?

Solution. A percentage is found by multiplying the relative frequency by 100%. Since the relative frequency for an interval is equal to the probability of falling in the interval, we must find the area to the right of $x = 35$ in Figure 7.16. This can be done by transforming $x = 35$ to the corresponding z value, obtaining the area to the left of z from Table A.4, and then subtracting this area from 1. We find that

$$z = \frac{35 - 30}{4.1} = 1.22.$$

Hence

$$\begin{aligned} P(X > 35) &= P(Z > 1.22) \\ &= 1 - P(Z < 1.22) \\ &= 1 - 0.8888 \\ &= 0.1112. \end{aligned}$$

Therefore, 11.12% of miniature poodles exceed 35 centimeters in height.

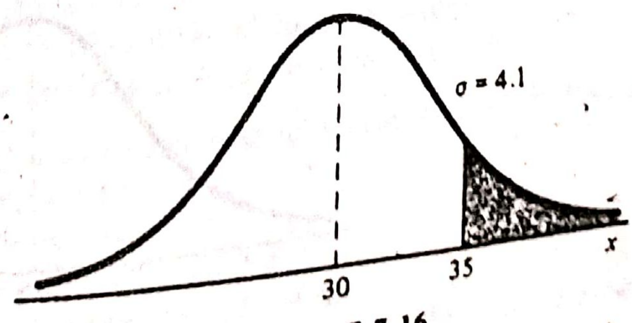

FIGURE 7.16
Area for Example 8.

Example 9. Find the percentage of miniature poodles exceeding 35 centimeters in Example 8 if the heights are all measured to the nearest centimeter.

Solution. This problem differs from Example 8 in that we now assign a measurement of 35 centimeters to all poodles whose heights are greater than 34.5 centimeters and less than 35.5 centimeters. We are actually approximating a discrete distribution by means of a continuous normal distribution. The required area is the region shaded to the right of 35.5 in Figure 7.17. We now find that

$$z = \frac{35.5 - 30}{4.1} = 1.34.$$

Hence

$$\begin{aligned} P(X > 35.5) &= P(Z > 1.34) \\ &= 1 - P(Z < 1.34) \\ &= 1 - 0.9099 \\ &= 0.0901. \end{aligned}$$

Therefore, 9.01% of miniature poodles exceed 35 centimeters in height when measured to the nearest centimeter. The difference of 2.11% between this answer and that of Example 8 represents all those poodles having a height greater than 35 and less than 35.5 centimeters that are now recorded as being 35 centimeters tall.

Example 10. The quality grade-point averages of 300 college freshmen follow approximately a normal distribution with a mean of 2.1 and a standard deviation of 0.8. How many of these freshmen would you expect to have a score between 2.5 and 3.5 inclusive if the point averages are computed to the nearest tenth?

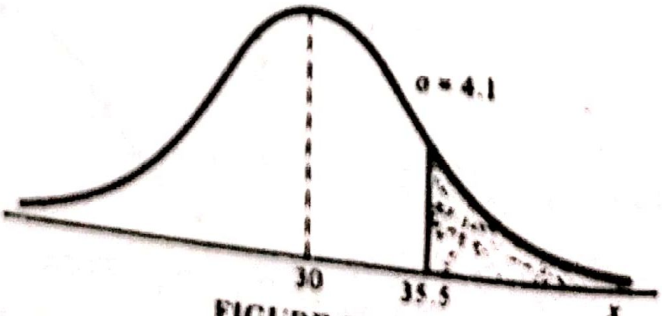

FIGURE 7.17
Area for Example 9.

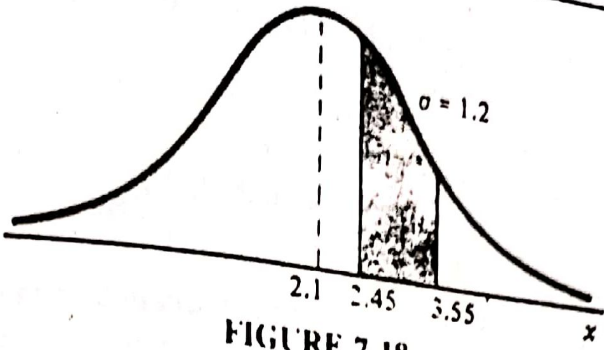

FIGURE 7.18
Area for Example 10.

Solution. Since the scores are recorded to the nearest tenth, we require the area between $x_1 = 2.45$ and $x_2 = 3.55$, as indicated in Figure 7.18. The corresponding z values are

$$z_1 = \frac{2.45 - 2.1}{0.8} = 0.44,$$

$$z_2 = \frac{3.55 - 2.1}{0.8} = 1.81.$$

Therefore,

$$\begin{aligned} P(2.45 < X < 3.55) &= P(0.44 < Z < 1.81) \\ &= P(Z < 1.81) - P(Z < 0.44) \\ &= 0.9649 - 0.6700 \\ &= 0.2949. \end{aligned}$$

Hence 29.49%, or approximately 88 of the 300 freshmen, should have a score between 2.5 and 3.5 inclusive.

1. Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find
   (a) the area below 32;
   (b) the area above 27;
   (c) the area between 42 and 51;
   (d) the x value that has 45% of the area below it;
   (e) the x value that has 13% of the area above it.
2. Given a normal distribution with $\mu = 200$ and $\sigma^2 = 100$, find
   (a) the area below 214;
   (b) the area above 179;
   (c) the area between 188 and 206;
   (d) the x value that has 80% of the area below it;
   (e) the two x values containing the middle 75% of the area.

3. Given the normally distributed variable X with mean 18 and standard deviation 2.5, find
   (a) $P(X < 15)$;
   (b) $P(17 < X < 21)$;
   (c) the value of k such that $P(X < k) = 0.2578$;
   (d) the value of k such that $P(X > k) = 0.1539$.

4. A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,
   (a) what fraction of the cups will contain more than 224 milliliters?
   (b) what is the probability that a cup contains between 191 and 209 milliliters?
   (c) how many cups will likely overflow if 230-milliliter cups are used for the next 1000 drinks?
   (d) below what value do we get the smallest 25% of the drinks?

5. The finished inside diameter of a piston ring is normally distributed with a mean of 10 centimeters and a standard deviation of 0.03 centimeter.
   (a) What proportion of rings will have inside diameters exceeding 10.075 centimeters?
   (b) What is the probability that a piston ring will have an inside diameter between 9.97 and 10.03 centimeters?
   (c) Below what value of inside diameter will 15% of the piston rings fall?

6. A lawyer commutes daily from his suburban home to his midtown office. On the average the trip one way takes 24 minutes, with a standard deviation of 3.8 minutes. Assume the distribution of trip times to be normally distributed.
   (a) What is the probability that a trip will take at least ½ hour?
   (b) If the office opens at 9:00 A.M. and he leaves his house at 8:45 A.M. daily, what percentage of the time is he late for work?
   (c) If he leaves the house at 8:35 A.M. and coffee is served at the office from 8:50 A.M. until 9:00 A.M., what is the probability that he misses coffee?
   (d) Find the length of time above which we find the slowest 15% of the trips.
   (e) Find the probability that 2 of the next 3 trips will take at least ½ hour.

7. If a set of grades on a statistics examination are approximately normally distributed with a mean of 74 and a standard deviation of 7.9, find
   (a) the lowest passing grade if the lowest 10% of the students are given F's;
   (b) the highest B if the top 5% of the students are given A's;
   (c) the lowest B if the top 10% of the students are given A's and the next 25% are given B's.

8. In a mathematics examination the average grade was 82 and the standard deviation was 5. All students with grades from 88 to 94 received a grade of B. If the grades are approximately normally distributed and 8 students received a B grade, how many students took the examination?

9. The heights of 1000 students are normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Assuming that the heights are recorded to the nearest half of a centimeter, how many of these students would you expect to have heights
   (a) less than 160.0 centimeters?
   (b) between 171.5 and 182.0 centimeters inclusive?
   (c) equal to 175.0 centimeters?
   (d) greater than or equal to 188.0 centimeters?

10. A company pays its employees an average wage of \$7.25 an hour with a standard deviation of 60 cents. If the wages are approximately normally distributed and paid to the nearest cent,
   (a) what percentage of the workers receive wages between \$6.75 and \$7.69 an hour inclusive?
   (b) the highest 5% of the employee hourly wages are greater than what amount?

11. The weights of a large number of miniature poodles are approximately normally distributed with a mean of 8 kilograms and a standard deviation of 0.9 kilogram. If measurements are recorded to the nearest tenth of a kilogram, find the fraction of these poodles with weights
   (a) over 9.5 kilograms;
   (b) at most 8.6 kilograms;
   (c) between 7.3 and 9.1 kilograms inclusive.

12. The tensile strength of a certain metal component is normally distributed with a mean of 10,000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter. Measurements are recorded to the nearest 50 kilograms per square centimeter.
   (a) What proportion of these components exceeds 10,150 kilograms per square centimeter in tensile strength?
   (b) If specifications require that all components have tensile strength between 9800 and 10,200 kilograms per square centimeter inclusive, what proportion of pieces would you expect to scrap?

13. If a set of observations is normally distributed, what percentage of the observations differs from the mean by
   (a) more than 1.3σ?
   (b) less than 0.52σ?

14. The IQs of 600 applicants to a certain college are approximately normally distributed with a mean of 115 and a standard deviation of 12. If the college requires an IQ of at least 95, how many of these students will be rejected on this basis regardless of their other qualifications?

15. The average rainfall, recorded to the nearest hundredth of a centimeter, in Roanoke, Virginia, for the month of March is 9.22 centimeters. Assuming a normal

distribution with a standard deviation of 2.83 centimeters, find the probability that next March Roanoke receives

(a) less than 1.84 centimeters of rain;

(b) more than 5 centimeters but not over 7 centimeters of rain;

(c) more than 13.8 centimeters of rain.

16. The average life of a certain type of small motor is 10 years, with a standard deviation of 2 years. The manufacturer replaces free all motors that fail while under guarantee. If he is willing to replace only 3% of the motors that fail, how long a guarantee should he offer? Assume that the lives of the motors follow a normal distribution.

7.4 Normal Approximation to the Binomial Distribution

Probabilities associated with binomial experiments are readily obtainable from the formula $b(x; n, p)$ of the binomial distribution or from Table A.2 when n is small. If n is not listed in any available table, we can compute the binomial probabilities by approximation procedures. In Section 6.5 we illustrated how the Poisson distribution can be used to approximate binomial probabilities when n is large and p is very close to zero or 1. Both the binomial and Poisson distributions are discrete. The first application of a continuous probability distribution to approximate probabilities over a discrete sample space was demonstrated in Section 7.3, Examples 9 and 10, where the normal curve was used. We now state a theorem that allows us to use areas under the normal curve to approximate binomial probabilities when n is sufficiently large.

THEOREM 7.1

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of

$$Z = \frac{X - np}{npq},$$

as $n \to \infty$, is the standardized normal distribution.

It turns out that the proper normal distribution provides a very accurate approximation to the binomial distribution when n is large and p is close to 1. In fact, even when n is small and p is not extremely close to zero or 1, the approximation is fairly good.

To investigate the normal approximation to the binomial distribution, we first draw the histogram for $b(x; 15, 0.4)$ and then superimpose the particular normal curve having the same mean and variance as the binomial variable X. Hence we draw a normal curve with

$$\mu = np = (15)(0.4) = 6,$$
$$\sigma^2 = npq = (15)(0.4)(0.6) = 3.6.$$

The histogram of $b(x; 15, 0.4)$ and the corresponding superimposed normal curve, which is completely determined by its mean and variance, are illustrated in Figure 7.19.

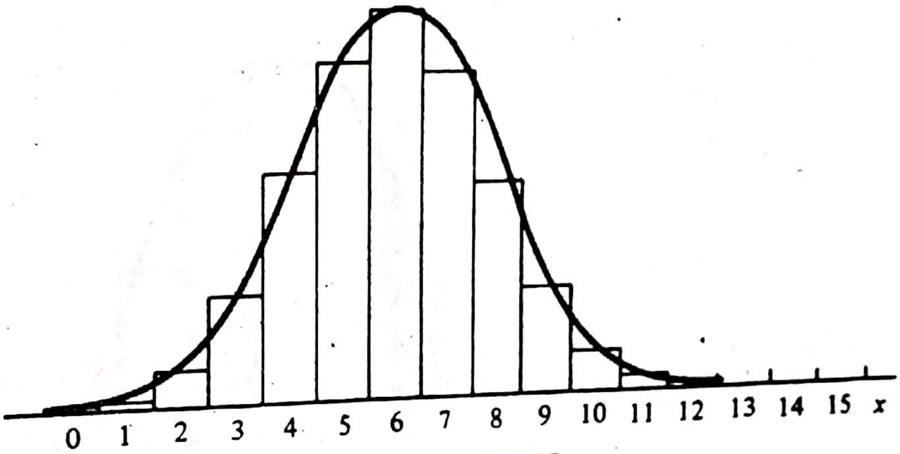

FIGURE 7.19
Normal-curve approximation of $b(x; 15, 0.4)$.

The exact probability of the binomial random variable X assuming a given value x is equal to the area of the rectangle whose base is centered at x. For example, the exact probability that X assumes the value 4 is equal to the area of the rectangle with base centered at $x = 4$. Using the formula for the binomial distribution, we find this area to be

$$b(4; 15, 0.4) = 0.1268.$$

This same probability is approximately equal to the area of the shaded region under the normal curve between the two ordinates $x_1 = 3.5$ and $x_2 = 4.5$ in Figure 7.20. Converting to z values, we have

$$z_1 = \frac{3.5 - 6}{1.9} = -1.316,$$

$$z_2 = \frac{4.5 - 6}{1.9} = -0.789.$$

If X is a binomial random variable and Z a standard normal variable, then

$$\begin{aligned} P(X = 4) &= b(4; 15, 0.4) \\ &\simeq P(-1.316 < Z < -0.789) \\ &= P(Z < -0.789) - P(Z < -1.316) \\ &= 0.2151 - 0.0941 \\ &= 0.1210. \end{aligned}$$

This agrees very closely with the exact value of 0.1268.

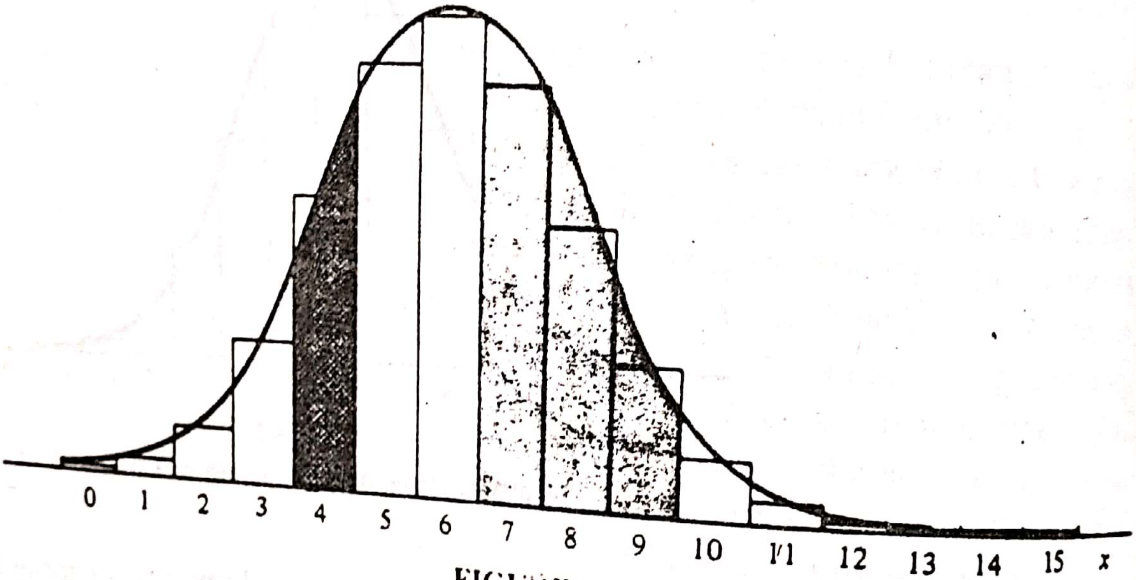

FIGURE 7.20
Normal approximation for $b(4; 15, 0.4)$ and $\sum_{x=7}^{9} b(x; 15, 0.4)$.

The normal approximation is most useful in calculating binomial sums for large values of n, which, without tables of binomial sums, is an impossible task. Referring to Figure 7.20, we might be interested in the probability that X assumes a value from 7 to 9 inclusive. The exact probability is given by

$$\begin{aligned} P(7 \leq X \leq 9) &= \sum_{x=7}^{9} b(x; 15, 0.4) \\ &= \sum_{x=0}^{9} b(x; 15, 0.4) - \sum_{x=0}^{6} b(x; 15, 0.4) \\ &= 0.9662 - 0.6098 \\ &= 0.3564, \end{aligned}$$

which is equal to the sum of the areas of the rectangles with bases centered at $x = 7$, 8, and 9. For the normal approximation we find the area of the shaded region under the curve between the ordinates $x_1 = 6.5$ and $x_2 = 9.5$ in Figure 7.20. The corresponding z values are

$$z_1 = \frac{6.5 - 6}{1.9} = 0.263,$$

$$z_2 = \frac{9.5 - 6}{1.9} = 1.842.$$

Now

$$\begin{aligned} P(7 \le X \le 9) &\simeq P(0.263 < Z < 1.842) \\ &= P(Z < 1.842) - P(Z < 0.263) \\ &= 0.9673 - 0.6037 \\ &= 0.3636. \end{aligned}$$

Once again the normal-curve approximation provides a value that agrees very closely with the exact value of 0.3564. The degree of accuracy, which depends on how well the curve fits the histogram, will increase as n increases. This is particularly true when p is not very close to $\frac{1}{2}$ and the histogram is no longer symmetric. Figures 7.21 and 7.22 show the histograms for $b(x; 6, 0.2)$ and $b(x; 15, 0.2)$, respectively. It is evident that a normal curve would fit the histogram when $n = 15$ considerably better than when $n = 6$.

In summary, we use the normal approximation to evaluate binomial prob-

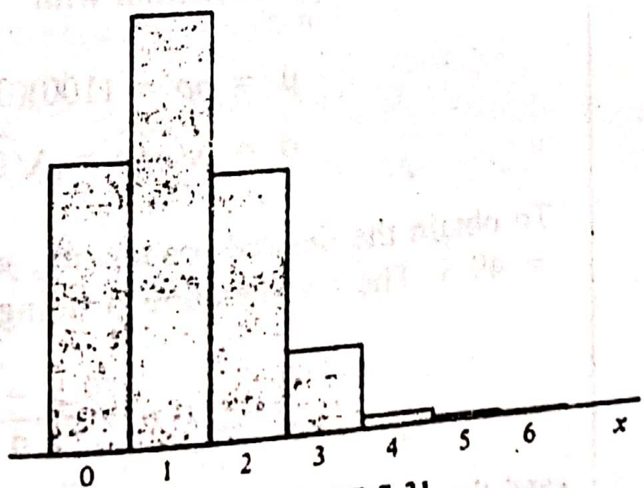

FIGURE 7.21
Histogram for $b(x; 6, 0.2)$

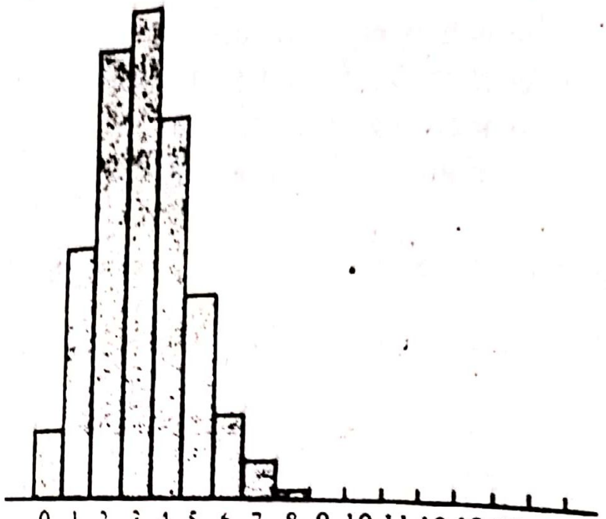

FIGURE 7.22
Histogram for $b(x; 15, 0.2)$.

abilities whenever p is not close to zero or 1. The approximation is excellent when n is large and fairly good for small values of n if p is reasonably close to $\frac{1}{2}$. One possible guide to determine when the normal approximation may be used is provided by calculating np and nq. If both np and nq are greater than 5, the approximation will be good.

Example 11. The probability that a patient recovers from a rare blood disease is 0.6. If 100 people are known to have contracted this disease, what is the probability that less than one-half survive?

Solution. Let the binomial variable X represent the number of patients that survive. Since $n = 100$, we should obtain fairly accurate results using the normal-curve approximation with

$$\mu = np = (100)(0.6) = 60,$$

$$\sigma = \sqrt{npq} = \sqrt{(100)(0.6)(0.4)} = 4.9.$$

To obtain the desired probability, we have to find the area to the left of $x = 49.5$. The z value corresponding to 49.5 is

$$z = \frac{49.5 - 60}{4.9} = -2.14,$$

and the probability of fewer than 50 of the 100 patients surviving is given

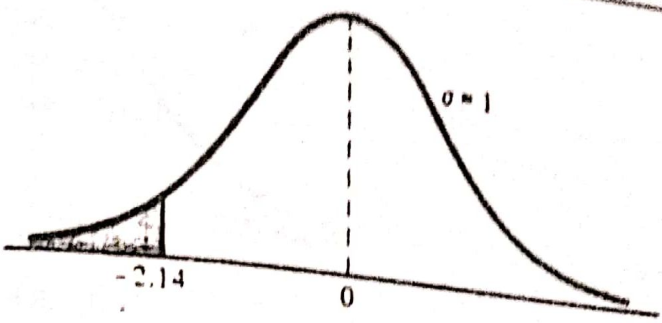

FIGURE 7.23
Area for Example 11.

by the area of the shaded region in Figure 7.23. Hence

$$\begin{aligned} P(X < 50) &= \sum_{x=0}^{49} b(x; 100, 0.6) \\ &\simeq P(Z < -2.14) \\ &= 0.0162. \end{aligned}$$

Example 12. A multiple-choice quiz has 200 questions, each with 4 possible answers, of which only 1 is the correct answer. What is the probability that sheer guesswork yields from 25 to 30 correct answers for 80 of the 200 problems about which the student has no knowledge?

Solution. The probability of a correct answer for each of the 80 questions is $p = \frac{1}{4}$. If X represents the number of correct answers due to guesswork, then

$$P(25 \le X \le 30) = \sum_{x=25}^{30} b(x; 80, \tfrac{1}{4}).$$

Using the normal-curve approximation with

$$\begin{aligned} \mu &= np = (80)(\tfrac{1}{4}) = 20, \\ \sigma &= \sqrt{npq} = \sqrt{(80)(\tfrac{1}{4})(\tfrac{3}{4})} = 3.87, \end{aligned}$$

we need the area between $x_1 = 24.5$ and $x_2 = 30.5$. The corresponding z values are

$$\begin{aligned} z_1 &= \frac{24.5 - 20}{3.87} = 1.16, \\ z_2 &= \frac{30.5 - 20}{3.87} = 2.71. \end{aligned}$$

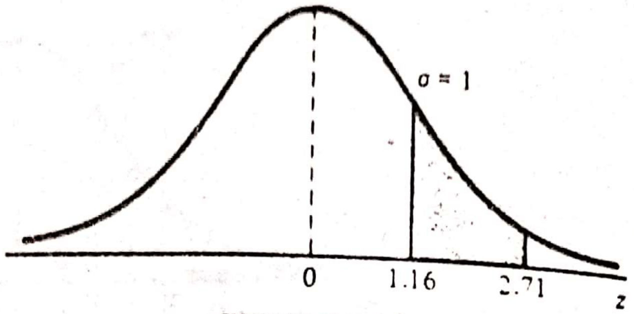

FIGURE 7.24
Area for Example 12.

The probability of correctly guessing from 25 to 30 questions is given by the area of the shaded region in Figure 7.24. From Table A.4 we find that

$$
\begin{aligned}
P(25 \leq X \leq 30) &= \sum_{x=25}^{30} b(x; 80, \tfrac{1}{4}) \\
&\simeq P(1.16 < Z < 2.71) \\
&= P(Z < 2.71) - P(Z < 1.16) \\
&= 0.9966 - 0.8770 \\
&= 0.1196.
\end{aligned}
$$

EXERCISES

1. Find the error in approximating $\sum_{x=1}^{4} b(x; 20, 0.1)$ by the normal-curve approximation
2. A coin is tossed 400 times. Use the normal-curve approximation to find the probability of obtaining
   (a) between 185 and 210 heads inclusive;
   (b) exactly 205 heads;
   (c) fewer than 176 or more than 227 heads.
3. A pair of dice is rolled 180 times. What is the probability that a total of 7 occurs
   (a) at least 25 times;
   (b) between 33 and 41 times inclusive;
   (c) exactly 30 times.
4. The probability that a patient recovers from a delicate heart operation is 0.9. Of the next 100 patients having this operation, what is the probability that
   (a) between 84 and 95 inclusive survive?
   (b) fewer than 86 survive?
5. A pheasant hunter claims that she brings down 75% of the birds she shoots at. Of the next 80 pheasants shot at by this hunter, what is the probability that
   (a) at least 50 escape?
   (b) at most 56 are brought down?

6. A drug manufacturer claims that a certain drug cures a blood disease on the average 80% of the time. To check the claim, government testers used the drug on a sample of 100 individuals and decide to accept the claim if 75 or more are cured.
   (a) What is the probability that the claim will be rejected when the cure probability is in fact 0.8?
   (b) What is the probability that the claim will be accepted by the government when the cure probability is as low as 0.7?

7. If 20% of the residents in a U.S. city prefer a white telephone over any other color available, what is the probability that among the next 1000 telephones installed in this city
   (a) between 170 and 185 inclusive will be white?
   (b) at least 210 but not more than 225 will be white?

8. One-sixth of the male freshmen entering a large state school are out-of-state students. If the students are assigned at random to the dormitories, 180 to a building, what is the probability that in a given dormitory at least one-fifth of the students are from out of state?

9. A certain pharmaceutical company knows that, on the average, 5% of a certain type of pill has an ingredient that is below the minimum strength and thus unacceptable. What is the probability that fewer than 10 in a sample of 200 pills will be unacceptable?

10. According to the May/June issue of *Consumers Digest*, census figures show that in 1978 almost 53% of all the households in the United States were composed of only one or two people. What is the probability that between 490 and 515 inclusive of the next 1000 randomly selected households in America will consist of either one or two people?