3. Select "Poisson" from the next menu.
4. Highlight the cells in which the values of x appear (i.e., cells A8 to A15).
5. Enter the mean number of arrivals (1.5).
6. Enter the logical "false" for the cumulative distribution.
7. Locate the cursor on cell B5. and copy the results to the remaining cells (i.e., cells B6 to B15).

The cumulative distribution listed in cells C8 to C15 is calculated in a similar fashion. However, in this case, enter the logical "true" in order to obtain the cumulative distribution.

Continuous Random Variables: The Normal Distribution

A second set of probability distributions should be applied when analyzing continuous random variables (i.e., variables that assume any value on the real number line). The statistical methods described in this text require an understanding of the normal distribution, the t distribution, the ^{2} (pronounced chi square) distribution, and the F distribution. The normal distribution is considered here. The t distribution and the F distribution are introduced in Chapter 6, and the ^{2} distribution is examined in Chapter 7.

The normal distribution is perhaps the most important of the theoretical distributions that are applied to continuous random variables. The following discussion focuses on the properties of the normal distribution, the use of the distribution in deriving probability, and the application of derived probabilities to problems or issues that occur in an administrative setting.

Exhibit 5.3 **An Application of the Poisson Probability Distribution**

Number of Emergent Cases x	Probability Distribution f(x)	Cumulative Probability Distribution F(x)
0	0.223	0.223
1	0.335	0.558
2	0.251	0.809
3	0.126	0.934
4	0.047	0.981
5	0.014	0.996
6	0.004	0.999
7	0.001	1.000

Rectangles that were constructed with equal width were used to form a bar chart (see Chapter 2). When a distribution depicts a relative frequency distribution and the width of the rectangles is equal to one, the area of each is interpreted as probability. For example, Exhibit 2.5 portrays the relative frequency distribution of patients, grouped by the number of secondary diagnoses. The area of the first rectangle indicates that if we select a patient at random the probability is 0.35 that the individual presents between one and four secondary conditions. Similarly, the probability of selecting a patient at random who presents 9 to 12 secondary conditions is 0.20. Consistent with the postulates of probability, the area of a given rectangle is interpreted as a probability; the probability ranges from zero to one, and the sum of the areas of the rectangles that comprise the bar chart is equal to one.

Figure 5.2 shows that the histogram constructed from the relative frequency distribution might be approximated by a smooth or continuous function. That the area *aegfd* of the continuous function is equivalent to area *abcd*, one of the rectangles forming the histogram, is demonstrated as follows. The area *aegcd* is common to both areas. Area *ebg* is equivalent to area *gcf*. Hence, area *aegfd* is equivalent to area *abcd*, and by extending this process, it can be demonstrated that the area formed by the continuous function is equivalent to the area that the histogram defined. Because the area formed by the histogram corresponds to probability, we conclude that the area under the continuous function also represents probability.

Similar to the smooth curve shown in Figure 5.2, the normal distribution is a continuous function and is perhaps the most important probability distribution that is used in inferential statistics. Figure 5.3 shows that the normal distribution is symmetrical, unimodal, and characterized by tails that are asymptotic with respect to the x axis. The property of symmetry implies that it is possible to con-

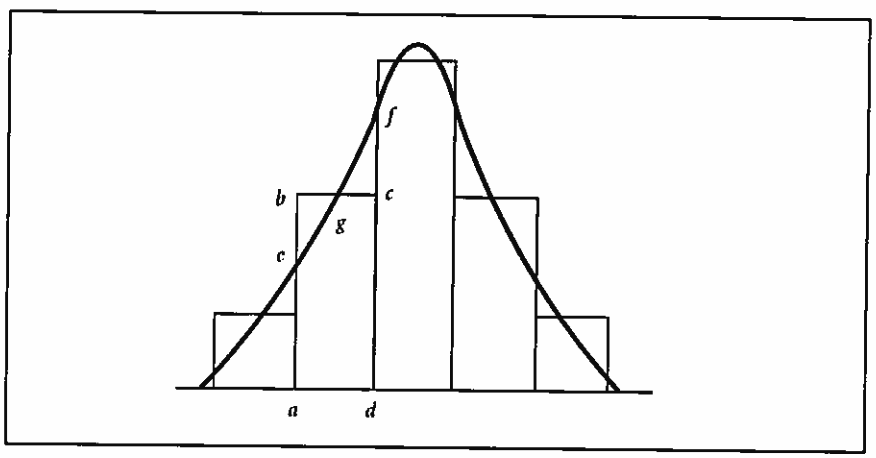

Figure 5.2 **Approximation of a Histogram by a Smooth Curve**

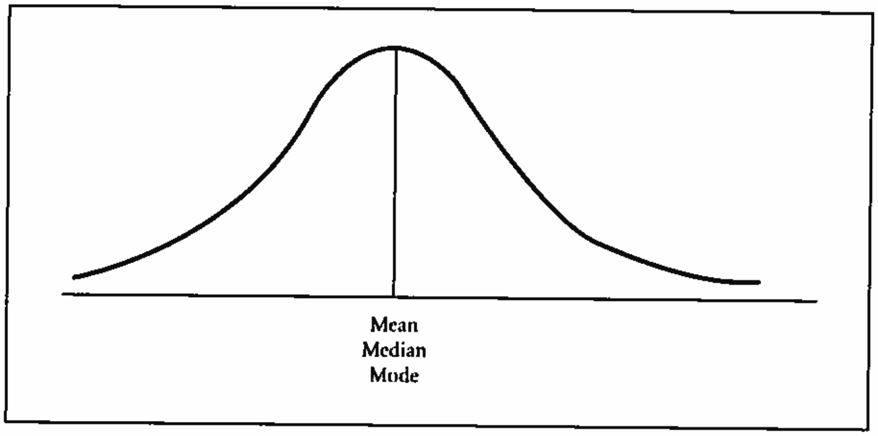

Figure 5.3 The Normal Distribution

struct a perpendicular line that divides the distribution into two identical halves (see Chapter 3). In addition, a symmetrical distribution is characterized by a mean, mode, and median that assume the same value.

Numerous distributions might be characterized as normal but differ in terms of scatter or variation and the value assumed by the arithmetic mean. For example, the distributions presented in Figure 5.4 are unimodal, symmetrical, and asymptotic with respect to the x axis. However, the two distributions differ in terms of the values assumed by the mean and the standard deviation. The probability density function for the normal distribution is determined by two parameters: the mean (μ) and the standard deviation (σ). In this case, the *probability density function* determines the height of the normal distribution relative to the x axis.

It is not possible to calculate the area under the probability density function for all possible values of μ and σ. Instead, after performing a scale transformation, we rely on the values that are presented in Table D.1 of Appendix D to determine the

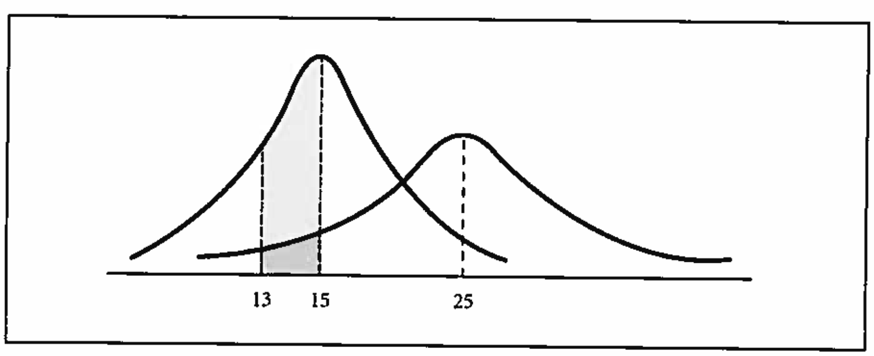

Figure 5.4 Two Different Normal Curves

area under a standard normal curve and, hence, probability. As before, a standard normal curve is unimodal, symmetrical, and asymptotic with respect to the x axis. However, the probability density function of the standard normal curve is defined specifically by a mean value of zero (i.e., $\mu = 0$) and a standard deviation of one (i.e., $\sigma = 1$).

The scale transformation to which we referred enables us to express a variable, measured in natural units, in terms of a standard normal deviate. Specifically, we define the standard normal deviate, represented by Z, by

$$Z = (x_i - \mu)/\sigma \qquad \text{(Equation 5.8)}$$

In this case, x_i represents the value of a continuous variable, measured in natural units. For example, suppose the administrator is assessing the staffing patterns in the operating theater and is interested in the time required to perform a given surgical procedure. Based on previous data, we assume that time required to perform the procedure is normally distributed with a mean time of 180 minutes and a standard deviation of 20 minutes. Suppose that the administrator is interested in an operative procedure with a duration of 220 minutes (Equation 5.8). In this case, we find that

$$Z = (220 - 180)/20$$

yields a standard normal deviate of 2.0. This suggests that an operative procedure lasting 220 minutes exceeds the mean time of 180 minutes by 2.0 standard deviations. Consider next a procedure that requires 150 minutes to complete. In this case, we find that

$$\begin{aligned} Z &= (150 - 180)/20 \\ &= -1.5 \end{aligned}$$

These calculations indicate that duration of the procedure is 1.5 standard deviations less than or below the average.

Equation 5.8 also might be used to transform a standard normal deviate into the original units of measure. Referring to the illustration involving the time required to perform the surgical procedure, suppose that the administrator is interested in the length of time that is 3 standard deviations in excess of the mean. In this case, we simply substitute appropriate values into Equation 5.8 and obtain

$$3.0 = (x_i - 180)/20$$

After rearranging slightly, we obtain

$$x_i = 3.0 \times 20 + 180$$

or 240 minutes. If we were interested in a procedure lasting 1 standard deviation less than the average, we would obtain

$$-1.0 = (x_i - 180)/20$$

After a slight rearrangement, we find that

$$x_i = -1.0 \times 20 + 180$$

or 160 minutes.

The results when Equation 5.8 is used allow us to determine probabilities from Table D.1. As indicated in Table D.1, values are presented for values of Z that range from Z = 0 to Z = 3.09. Suppose, for example, that we wished to determine and interpret the value for a standard normal deviate equal to 1.96. In this case, the desired value appears in the cell that corresponds to the intersection of the row identified by the value 1.9 and the column identified by the value 0.06 (i.e., the sum of 1.9 and 0.06 is the desired standard normal deviate). After locating the cell identified by the intersection, we observe that the value is 0.4750, implying that 47.5% of the area under the probability density function ranges from Z = 0 to Z = 1.96. Thus, we conclude that the probability of obtaining a standard normal deviate that is greater than or equal to zero or less than or equal to 1.96 is 0.4750. This result might be expressed as

$$P(0 \leq Z \leq 1.96) = 0.4750$$

The probability of obtaining a standard normal deviate in excess of 1.96 is 0.025 or less (i.e., the difference between 0.5000 and 0.4750).

For example, if the administrator is interested in the probability that an operative procedure will last more than 210 minutes, the corresponding standard normal deviate is given by

$$Z = (210 - 180)/20$$

or 1.5. Table D.1 shows that a standard normal deviate of 1.5 is 0.4332, implying that the probability of a given operative procedure lasting more than 210 minutes is 0.5000 − 0.4332. These calculations indicate that the likelihood of the operative procedure requiring more than 210 minutes is 0.0668 or less.

It might be interesting to determine the probability that an operative procedure will last between 190 and 220 minutes. In this case, the standard normal deviate corresponding to 190 minutes is given by

$$Z_{190} = (190 - 180)/20$$

or 0.5, whereas the standard normal deviate corresponding to 220 minutes is given by

$$Z_{200} = (220 - 180)/20$$

or 2.0. The value of Z_{190} or 0.5 is 0.1915. On the other hand, the tabular value of Z_{200} or 2.0 is 0.4772. Thus, consistent with Figure 5.5, we assert that

$$\begin{aligned} P(0 \leq Z \leq 2.0)&: \quad 0.4772 \\ P(0 \leq Z \leq 0.5)&: \quad -\underline{0.1915} \\ P(0.5 < Z < 2.0) &= 0.2875 \end{aligned}$$

Thus, the probability of the procedure lasting between 190 and 220 minutes is approximately 0.29.

As a final example, if the administrator wishes to determine the probability that a given procedure will last between 140 and 250 minutes, we transform 140 and 210 minutes into standard normal deviates as follows:

$$\begin{aligned} Z_{140} &= (140 - 180)/20 \\ &= -2.0 \\ Z_{210} &= (210 - 180)/20 \\ &= 1.5 \end{aligned}$$

Regarding the finding that $Z_{140} = -2.0$, the standard normal curve is symmetrical, implying that the probability $P(0 \leq Z \leq 2.0)$ is identical to $P(-2.0 \leq Z \leq 0)$, and as a consequence, no negative values for the standard normal deviate appear in the

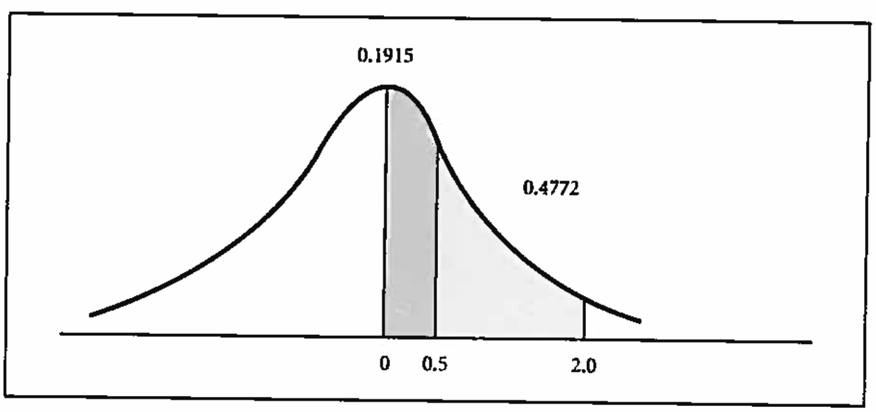

Figure 5.5 **The P (0.5 < Z < 2.0)**

table. Because $P(0 \leq Z \leq 2.0)$ and $P(-2.0 \leq Z \leq 0)$ are the same, we simply substitute the value that corresponds to $Z = 2$ for the value of $Z = -2$ and obtain

$$P(-2.0 \leq Z \leq 0) = 0.4772$$

The table value for a standard normal deviate of 1.5 is 0.4332, and Figure 5.6 shows that the desired probability is given by the sum

$$\begin{array}{ll} P(-2.0 \leq Z \leq 0): & 0.4772 \\ P(0 \leq Z \leq 1.5): & \underline{0.4332} \\ P(-2.0 \leq Z \leq 1.5): & 0.9104 \end{array}$$

These results indicate that the probability of an operative procedure lasting between 140 and 210 minutes is approximately 0.91.

Sampling Distributions

We now turn attention to perhaps the most important of the concepts in inferential statistics: the sampling distribution of a statistic, a concept that emphasizes chance variation and the role of dispersion in inferential statistics. It is possible to identify essentially two approaches to the study of sampling distributions. The first results in an experimental or an observed sampling distribution that is based on the data obtained by selecting a series of repeated samples from a given population. The second approach relies on mathematical theory and results in a theoretical sampling distribution of the statistic. Each of the two approaches is discussed here.

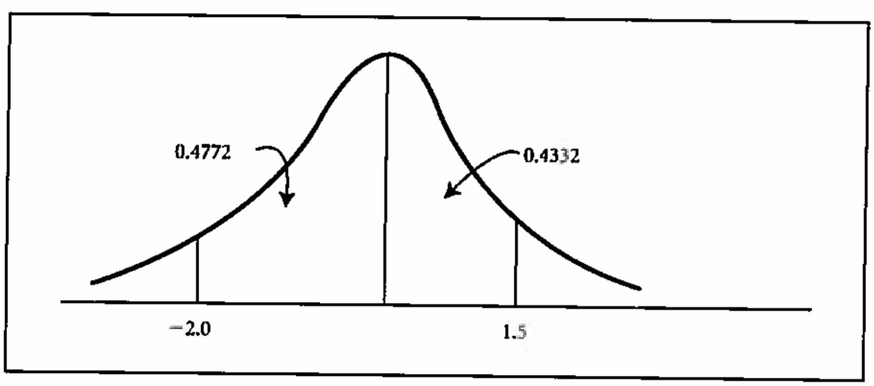

Figure 5.6 **Finding the Probability P (−2.0 ≤ Z ≤ 1.5)**