CHAPTER 5

Discrete Distributions

LEARNING OBJECTIVES

The overall learning objective of Chapter 5 is to help you understand a category of probability distributions that produces only discrete outcomes, thereby enabling you to:

1. Define a random variable in order to differentiate between a discrete distribution and a continuous distribution
2. Determine the mean, variance, and standard deviation of a discrete distribution
3. Solve problems involving the binomial distribution using the binomial formula and the binomial table
4. Solve problems involving the Poisson distribution using the Poisson formula and the Poisson table
5. Solve problems involving the hypergeometric distribution using the hypergeometric formula

Noel Hendrickson/DigitalVision/Getty Images

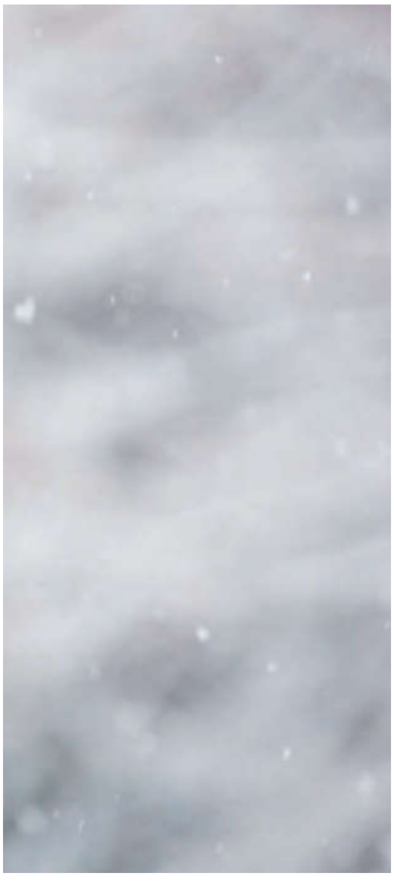
Life with a Cell Phone

DECISION DILEMMA

Decision Dilemma

As early as 1947, scientists understood the basic concept of a cell phone as a type of two-way radio. Seeing the potential of crude mobile car phones, researchers understood that by using a small range of service areas (cells) with frequency reuse, they could increase the capacity for mobile phone usage significantly even though the technology was not then available. During that same year, AT&T proposed the allocation of a large number of radio-spectrum frequencies by the FCC that would thereby make widespread mobile phone service feasible. At the same time, the FCC decided to limit the amount of frequency capacity available such that only 23 phone conversations could take place simutaneously. In 1968, the FCC reconsidered its position and freed the airwaves for more phones. About this time, AT&T and Bell Labs proposed to the FCC a system in which they would construct a series of many small, low-powered broadcast towers, each of which would broadcast to a "cell" covering a few miles. Taken as a whole, such "cells" could be used to pass phone calls from cell to cell, thereby reaching a large area.

The first company to actually produce a cell phone was Motorola, and Dr. Martin Cooper, then of Motorola and considered the inventor of the first modern portable handset, made his first call on the portable cell phone in 1973. By 1977, AT&T and Bell Labs had developed a prototype cellular phone system that was tested in Chicago by 2,000 trial customers. After the first commercial cell phone system began operation in Japan in 1979, and Motorola and American Radio developed a second U.S. cell system in 1981, the FCC authorized commerical cellular service in the United States in 1982. By 1987, cell phone subscribers had exceeded 1 million customers in the United States, and as frequencies were getting crowded, the FCC authorized alternative cellular technologies, opening up new opportunities for development. Since that time, researchers have developed a number of advances that have increased capacity exponentially.



Today in the United States, over 14% of cell phone owners use only cellular phones, and the trend is rising. According to a Harris Poll of 9132 surveyed adults, 89% of adults have a cell phone. In an Associated Press/America Online Pew Poll of 1,200 cell phone users, it was discovered that two-thirds of all cell phone users said that it would be hard to give up their cell phones, and 26% responded that they cannot imagine life without their cell phones. In spite of American's growing dependence on their cell phones, not everyone is happy about their usage. Almost 9 out of 10 cell users encounter others using their phones in an annoying way. In addition, 28% claim that sometimes they do not drive as safely as they should because they are using cell phones. Now, there are multiple uses for the cell phone, including picture taking, text messaging, game playing, and others. According to the study, two-thirds of cell phone owners in the 18 to 29 age bracket sent text messages using their cell phones, 55% take pictures with their phones, 47% play games on the phones, and 28% use the Internet through their cell phones.

Managerial and Statistical Questions

1. One study reports that 14% of cell phone owners in the United States use only cellular phones (no land line). Suppose you randomly select 20 Americans, what is the probability that more than 7 of the sample use only cell phones?
2. The study also reports that 9 out of 10 cell users encounter others using their phones in an annoying way. Based on this, if you were to randomly select 25 cell phone users, what is the probability that fewer than 20 report that they encounter others using their phones in an annoying way?
3. Suppose a survey of cell phone users shows that, on average, a cell phone user receives 3.6 calls per day. If this figure is true, what is the probability that a cell phone user receives no calls in a day? What is the probability that a cell phone user receives five or more calls in a day?

Sources: Mary Bellis, "Selling the Cell Phone, Part 1: History of Cellular Phones," in *About Business & Finance.* An America Online site, Selling the Cell Phone—History of Cellular Phones at: http://inventors.about.com/library/weekly/aa070899.htm; *USA Today* Tech, "For Many, Their Cell Phone Has Become Their Only Phone," at: http://www.usatoday.com/tech/news/2003-03-24-cell-phones x.htm; and Will Lester, "A Love-Hate Relationship," *Houston Chronicle.* April 4, 2006, p. D4. http://www.harrisinteractive.com/harris_poll/index.asp?PID=890

TABLE 5.1

All Possible Outcomes for the Battery Experiment

G_1	G_2	G_3
D_1	G_2	G_3
G_1	D_2	G_3
G_1	G_2	D_3
D_1	D_2	G_3
D_1	G_2	D_3
G_1	D_2	D_3
D_1	D_2	D_3

In statistical experiments involving chance, outcomes occur randomly. As an example of such an experiment, a battery manufacturer randomly selects three batteries from a large batch of batteries to be tested for quality. Each selected battery is to be rated as good or defective. The batteries are numbered from 1 to 3, a defective battery is designated with a D, and a good battery is designated with a G. All possible outcomes are shown in Table 5.1. The expression $D_1\ G_2\ D_3$ denotes one particular outcome in which the first and third batteries are defective and the second battery is good. In this chapter, we examine the probabilities of events occurring in experiments that produce discrete distributions. In particular, we will study the binomial distribution, the Poisson distribution, and the hypergeometric distribution.

5.1 DISCRETE VERSUS CONTINUOUS DISTRIBUTIONS

A **random variable** is *a variable that contains the outcomes of a chance experiment.* For example, suppose an experiment is to measure the arrivals of automobiles at a turnpike tollbooth during a 30-second period. The possible outcomes are: 0 cars, 1 car, 2 cars, . . . , n cars. These numbers (0, 1, 2, . . . , n) are the values of a random variable. Suppose another experiment is to measure the time between the completion of two tasks in a production line. The values will range from 0 seconds to n seconds. These time measurements are the values of another random variable. The two categories of random variables are (1) discrete random variables and (2) continuous random variables.

A random variable is a **discrete random variable** *if the set of all possible values is at most a finite or a countably infinite number of possible values.* In most statistical situations, discrete random variables produce values that are nonnegative whole numbers. For example, if six people are randomly selected from a population and how many of the six are left-handed is to be determined, the random variable produced is discrete. The only possible numbers of left-handed people in the sample of six are 0, 1, 2, 3, 4, 5, and 6. There cannot be 2.75 left-handed people in a group of six people; obtaining nonwhole number values is impossible. Other examples of experiments that yield discrete random variables include the following:

1. Randomly selecting 25 people who consume soft drinks and determining how many people prefer diet soft drinks
2. Determining the number of defects in a batch of 50 items
3. Counting the number of people who arrive at a store during a five-minute period
4. Sampling 100 registered voters and determining how many voted for the president in the last election

The battery experiment described at the beginning of the chapter produces a distribution that has discrete outcomes. Any one trial of the experiment will contain 0, 1, 2, or 3 defective batteries. It is not possible to get 1.58 defective batteries. It could be said that discrete random variables are usually generated from experiments in which things are "counted" not "measured."

Continuous random variables *take on values at every point over a given interval.* Thus continuous random variables have no gaps or unassumed values. It could be said that continuous random variables are generated from experiments in which things are "measured" not "counted." For example, if a person is assembling a product component, the time it takes to accomplish that feat could be any value within a reasonable range such as 3 minutes 36.4218 seconds or 5 minutes 17.5169 seconds. A list of measures for which continuous random variables might be generated would include time, height, weight, and volume. Other examples of experiments that yield continuous random variables include the following:

1. Sampling the volume of liquid nitrogen in a storage tank
2. Measuring the time between customer arrivals at a retail outlet
3. Measuring the lengths of newly designed automobiles
4. Measuring the weight of grain in a grain elevator at different points of time

Once continuous data are measured and recorded, they become discrete data because the data are rounded off to a discrete number. Thus in actual practice, virtually all business data are discrete. However, for practical reasons, data analysis is facilitated greatly by using continuous distributions on data that were continuous originally.

The outcomes for random variables and their associated probabilities can be organized into distributions. The two types of distributions are **discrete distributions**, *constructed from discrete random variables,* and **continuous distributions**, *based on continuous random variables.*

In this text, three discrete distributions are presented:

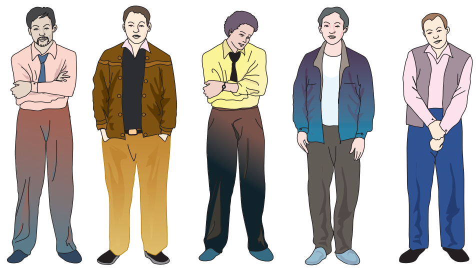
1. binomial distribution
2. Poisson distribution
3. hypergeometric distribution

All three of these distributions are presented in this chapter.

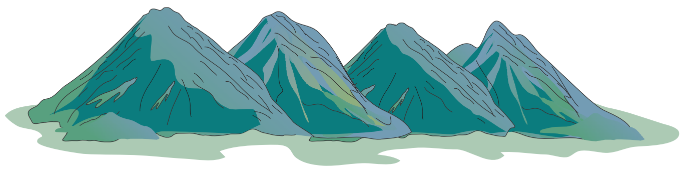
In addition, six continuous distributions are discussed later in this text:

1. uniform distribution
2. normal distribution
3. exponential distribution
4. *t* distribution
5. chi-square distribution
6. *F* distribution

Discrete



Continuous



5.2 DESCRIBING A DISCRETE DISTRIBUTION

How can we describe a discrete distribution? One way is to construct a graph of the distribution and study the graph. The histogram is probably the most common graphical way to depict a discrete distribution.

Observe the discrete distribution in Table 5.2. An executive is considering out-of-town business travel for a given Friday. She recognizes that at least one crisis could occur on the day that she is gone and she is concerned about that possibility. Table 5.2 shows a discrete distribution that contains the number of crises that could occur during the day that she is gone and the probability that each number will occur. For example, there is a .37 probability that no crisis will occur, a .31 probability of one crisis, and so on. The histogram in Figure 5.1 depicts the distribution given in Table 5.2. Notice that the *x*-axis of the histogram contains the possible outcomes of the experiment (number of crises that might occur) and that the *y*-axis contains the probabilities of these occurring.

It is readily apparent from studying the graph of Figure 5.1 that the most likely number of crises is 0 or 1. In addition, we can see that the distribution is discrete in that no probabilities are shown for values in between the whole-number crises.

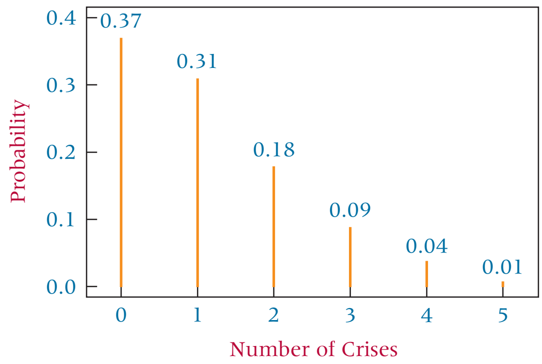
TABLE 5.2

Discrete Distribution of Occurrence of Daily Crises

Number of Crises	Probability
0	.37
1	.31
2	.18
3	.09
4	.04
5	.01

FIGURE 5.1

Minitab Histogram of Discrete Distribution of Crises Data



Mean, Variance, and Standard Deviation of Discrete Distributions

What additional mechanisms can be used to describe discrete distributions besides depicting them graphically? The measures of central tendency and measures of variability discussed in Chapter 3 for grouped data can be applied to discrete distributions to compute a mean, a variance, and a standard deviation. Each of those three descriptive measures (mean, variance, and standard deviation) is computed on grouped data by using the class midpoint as the value to represent the data in the class interval. With discrete distributions, using the class midpoint is not necessary because the discrete value of an outcome (0, 1, 2, 3, . . .) is used to represent itself. Thus, instead of using the value of the class midpoint (M) in computing these descriptive measures for grouped data, the discrete experiment's outcomes (x) are used. In computing these descriptive measures on grouped data, the frequency of each class interval is used to weight the class midpoint. With discrete distribution analysis, the probability of each occurrence is used as the weight.

TABLE 5.3

Computing the Mean of the Crises Data

x	$P(x)$	$x \cdot P(x)$
0	.37	.00
1	.31	.31
2	.18	.36
3	.09	.27
4	.04	.16
5	.01	.05
		$\Sigma[x \cdot P(x)] = 1.15$
		$\mu = 1.15$ crises

Mean or Expected Value

The **mean** or **expected value** of a discrete distribution is *the long-run average of occurrences*. We must realize that any one trial using a discrete random variable yields only one outcome. However, if the process is repeated long enough, the average of the outcomes are most likely to approach a long-run average, expected value, or mean value. This mean, or expected, value is computed as follows.

MEAN OR EXPECTED VALUE OF A DISCRETE DISTRIBUTION

$$\mu = E(x) = \Sigma[x \cdot P(x)]$$

where

$E(x)$ = long-run average
x = an outcome
$P(x)$ = probability of that outcome

As an example, let's compute the mean or expected value of the distribution given in Table 5.2. See Table 5.3 for the resulting values. In the long run, the mean or expected number of crises on a given Friday for this executive is 1.15 crises. Of course, the executive will never have 1.15 crises.

Variance and Standard Deviation of a Discrete Distribution

The variance and standard deviation of a discrete distribution are solved for by using the outcomes (x) and probabilities of outcomes [$P(x)$] in a manner similar to that of computing a

TABLE 5.4

Calculation of Variance and Standard Deviation on Crises Data

x	$P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
0	.37	$(0-1.15)^2 = 1.32$	$(1.32)(.37) = .49$
1	.31	$(1-1.15)^2 = .02$	$(0.02)(.31) = .01$
2	.18	$(2-1.15)^2 = .72$	$(0.72)(.18) = .13$
3	.09	$(3-1.15)^2 = 3.42$	$(3.42)(.09) = .31$
4	.04	$(4-1.15)^2 = 8.12$	$(8.12)(.04) = .32$
5	.01	$(5-1.15)^2 = 14.82$	$(14.82)(.01) = .15$
			$\Sigma[(x-\mu)^2 \cdot P(x)] = 1.41$

The variance of $\sigma^2 = \Sigma[(x-\mu)^2 \cdot P(x)] = 1.41$

The standard deviation is $\sigma = \sqrt{1.41} = 1.19$ crises.

mean. In addition, the computations for variance and standard deviations use the mean of the discrete distribution. The formula for computing the variance follows.

VARIANCE OF A DISCRETE DISTRIBUTION

$$\sigma^2 = \Sigma[(x-\mu)^2 \cdot P(x)]$$

where

x = an outcome
$P(x)$ = probability of a given outcome
μ = mean

The standard deviation is then computed by taking the square root of the variance.

STANDARD DEVIATION OF A DISCRETE DISTRIBUTION

$$\sigma = \sqrt{\Sigma[(x-\mu)^2 \cdot P(x)]}$$

The variance and standard deviation of the crisis data in Table 5.2 are calculated and shown in Table 5.4. The mean of the crisis data is 1.15 crises. The standard deviation is 1.19 crises, and the variance is 1.41.

DEMONSTRATION PROBLEM 5.1

During one holiday season, the Texas lottery played a game called the Stocking Stuffer. With this game, total instant winnings of $34.8 million were available in 70 million $1 tickets, with ticket prizes ranging from $1 to $1,000. Shown here are the various prizes and the probability of winning each prize. Use these data to compute the expected value of the game, the variance of the game, and the standard deviation of the game.

Prize (x)	Probability $P(x)$
$1,000	.00002
100	.00063
20	.00400
10	.00601
4	.02403
2	.08877
1	.10479
0	.77175

Solution

The mean is computed as follows.

Prize (x)	Probability $P(x)$	$x \cdot P(x)$
$1,000	.00002	.02000
100	.00063	.06300
20	.00400	.08000
10	.00601	.06010
4	.02403	.09612
2	.08877	.17754
1	.10479	.10479
0	.77175	.00000
		$\Sigma[x \cdot P(x)] = .60155$

$$\mu = E(x) = \Sigma[x \cdot P(x)] = .60155$$

The expected payoff for a $1 ticket in this game is 60.2 cents. If a person plays the game for a long time, he or she could expect to average about 60 cents in winnings. In the long run, the participant will lose about $1.00 − .602 = .398, or about 40 cents a game. Of course, an individual will never win 60 cents in any one game.

Using this mean, $\mu = .60155$, the variance and standard deviation can be computed as follows.

x	$P(x)$	$(x - \mu)^2$	$(x - \mu)^2 \cdot P(x)$
$1,000	.00002	998797.26190	19.97595
100	.00063	9880.05186	6.22443
20	.00400	376.29986	1.50520
10	.00601	88.33086	0.53087
4	.02403	11.54946	0.27753
2	.08877	1.95566	0.17360
1	.10479	0.15876	0.01664
0	.77175	0.36186	0.27927
			$\Sigma[(x - \mu)^2 \cdot P(x)] = 28.98349$

$$\sigma^2 = \Sigma[(x - \mu)^2 \cdot P(x)] = 28.98349$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\Sigma[(x - \mu)^2 \cdot P(x)]} = \sqrt{28.98349} = 5.38363$$

The variance is 28.98349 (dollars)^2 and the standard deviation is $5.38.

5.2 PROBLEMS

5.1 Determine the mean, the variance, and the standard deviation of the following discrete distribution.

x	$P(x)$
1	.238
2	.290
3	.177
4	.158
5	.137

5.2 Determine the mean, the variance, and the standard deviation of the following discrete distribution.

x	$P(x)$
0	.103
1	.118
2	.246
3	.229
4	.138
5	.094
6	.071
7	.001

5.3 The following data are the result of a historical study of the number of flaws found in a porcelain cup produced by a manufacturing firm. Use these data and the associated probabilities to compute the expected number of flaws and the standard deviation of flaws.

Flaws	Probability
0	.461
1	.285
2	.129
3	.087
4	.038

5.4 Suppose 20% of the people in a city prefer Pepsi-Cola as their soft drink of choice. If a random sample of six people is chosen, the number of Pepsi drinkers could range from zero to six. Shown here are the possible numbers of Pepsi drinkers in a sample of six people and the probability of that number of Pepsi drinkers occurring in the sample. Use the data to determine the mean number of Pepsi drinkers in a sample of six people in the city, and compute the standard deviation.

Number of Pepsi Drinkers	Probability
0	.262
1	.393
2	.246
3	.082
4	.015
5	.002
6	.000

5.3 BINOMIAL DISTRIBUTION

Video

Interactive Applet

Perhaps the most widely known of all discrete distributions is the **binomial distribution**. The binomial distribution has been used for hundreds of years. Several assumptions underlie the use of the binomial distribution:

ASSUMPTIONS OF THE BINOMIAL DISTRIBUTION

- The experiment involves n identical trials.
- Each trial has only two possible outcomes denoted as success or as failure.
- Each trial is independent of the previous trials.
- The terms p and q remain constant throughout the experiment, where the term p is the probability of getting a success on any one trial and the term $q = (1 - p)$ is the probability of getting a failure on any one trial.

As the word *binomial* indicates, any single trial of a binomial experiment contains only two possible outcomes. These two outcomes are labeled *success* or *failure*. Usually the outcome of interest to the researcher is labeled a success. For example, if a quality control analyst is looking for defective products, he would consider finding a defective product a success even though the company would not consider a defective product a success. If researchers are studying left-handedness, the outcome of getting a left-handed person in a trial of an experiment is a success. The other possible outcome of a trial in a binomial experiment is called a failure. The word *failure* is used only in opposition to success. In the preceding experiments, a failure could be to get an acceptable part (as opposed to a defective part) or to get a right-handed person (as opposed to a left-handed person). In a binomial distribution experiment, any one trial can have only two possible, mutually exclusive outcomes (right-handed/left-handed, defective/good, male/female, etc.).

The binomial distribution is a discrete distribution. In n trials, only x successes are possible, where x is a whole number between 0 and n. For example, if five parts are randomly selected from a batch of parts, only 0, 1, 2, 3, 4, or 5 defective parts are possible in that sample. In a sample of five parts, getting 2.714 defective parts is not possible, nor is getting eight defective parts possible.

In a binomial experiment, the trials must be independent. This constraint means that either the experiment is by nature one that produces independent trials (such as tossing coins or rolling dice) or the experiment is conducted with replacement. The effect of the independent trial requirement is that p, the probability of getting a success on one trial, remains constant from trial to trial. For example, suppose 5% of all parts in a bin are defective. The probability of drawing a defective part on the first draw is $p = .05$. If the first part drawn is not replaced, the second draw is not independent of the first, and the p value will change for the next draw. The binomial distribution does not allow for p to change from trial to trial within an experiment. However, if the population is large in comparison with the sample size, the effect of sampling without replacement is minimal, and the independence assumption essentially is met, that is, p remains relatively constant.

Generally, if the sample size, n, is less than 5% of the population, the independence assumption is not of great concern. Therefore the acceptable sample size for using the binomial distribution with samples taken *without* replacement is

$$n < 5\%N$$

where

n = sample size
N = population size

For example, suppose 10% of the population of the world is left-handed and that a sample of 20 people is selected randomly from the world's population. If the first person selected is left-handed—and the sampling is conducted without replacement—the value of $p = .10$ is virtually unaffected because the population of the world is so large. In addition, with many experiments the population is continually being replenished even as the sampling is being done. This condition often is the case with quality control sampling of products from large production runs. Some examples of binomial distribution problems follow.

1. Suppose a machine producing computer chips has a 6% defective rate. If a company purchases 30 of these chips, what is the probability that none is defective?
2. One ethics study suggested that 84% of U.S. companies have an ethics code. From a random sample of 15 companies, what is the probability that at least 10 have an ethics code?
3. A survey found that nearly 67% of company buyers stated that their company had programs for preferred buyers. If a random sample of 50 company buyers is taken, what is the probability that 40 or more have companies with programs for preferred buyers?

Solving a Binomial Problem

A survey of relocation administrators by Runzheimer International revealed several reasons why workers reject relocation offers. Included in the list were family considerations, financial reasons, and others. Four percent of the respondents said they rejected relocation offers because they received too little relocation help. Suppose five workers who just rejected relocation offers are randomly selected and interviewed. Assuming the 4% figure holds for all workers rejecting relocation, what is the probability that the first worker interviewed rejected the offer because of too little relocation help and the next four workers rejected the offer for other reasons?

Let T represent too little relocation help and R represent other reasons. The sequence of interviews for this problem is as follows:

$$T_1, R_2, R_3, R_4, R_5$$

The probability of getting this sequence of workers is calculated by using the special rule of multiplication for independent events (assuming the workers are independently selected from a large population of workers). If 4% of the workers rejecting relocation offers do so for too little relocation help, the probability of one person being randomly selected from workers rejecting relocation offers who does so for that reason is .04, which is the value of p. The other 96% of the workers who reject relocation offers do so for other reasons. Thus the probability of randomly selecting a worker from those who reject relocation offers who does so for other reasons is $1 - .04 = .96$, which is the value for q. The probability of obtaining this sequence of five workers who have rejected relocation offers is

$$P(T_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5) = (.04)(.96)(.96)(.96)(.96) = .03397$$

Obviously, in the random selection of workers who rejected relocation offers, the worker who did so because of too little relocation help could have been the second worker or the third or the fourth or the fifth. All the possible sequences of getting one worker who rejected relocation because of too little help and four workers who did so for other reasons follow.

$$T_1, R_2, R_3, R_4, R_5$$
$$R_1, T_2, R_3, R_4, R_5$$
$$R_1, R_2, T_3, R_4, R_5$$
$$R_1, R_2, R_3, T_4, R_5$$
$$R_1, R_2, R_3, R_4, T_5$$

The probability of each of these sequences occurring is calculated as follows:

$$(.04)(.96)(.96)(.96)(.96) = .03397$$
$$(.96)(.04)(.96)(.96)(.96) = .03397$$
$$(.96)(.96)(.04)(.96)(.96) = .03397$$
$$(.96)(.96)(.96)(.04)(.96) = .03397$$
$$(.96)(.96)(.96)(.96)(.04) = .03397$$

Note that in each case the final probability is the same. Each of the five sequences contains the product of .04 and four .96s. The commutative property of multiplication allows for the reordering of the five individual probabilities in any one sequence. The probabilities in each of the five sequences may be reordered and summarized as $(.04)^1\,(.96)^4$. Each sequence contains the same five probabilities, which makes recomputing the probability of each sequence unnecessary. What *is* important is to determine how many different ways the sequences can be formed and multiply that figure by the probability of one sequence occurring. For the five sequences of this problem, the total probability of getting exactly one worker who rejected relocation because of too little relocation help in a random sample of five workers who rejected relocation offers is

$$5(.04)^1(.96)^4 = .16987$$

An easier way to determine the number of sequences than by listing all possibilities is to use *combinations* to calculate them. (The concept of combinations was introduced in Chapter 4.) Five workers are being sampled, so $n = 5$, and the problem is to get one worker who rejected a relocation offer because of too little relocation help, $x = 1$. Hence ${}_nC_x$ will yield the number of possible ways to get x successes in n trials. For this problem, ${}_5C_1$ tells the number of sequences of possibilities.

$${}_5C_1 = \frac{5!}{1!(5-1)!} = 5$$

Weighting the probability of one sequence with the combination yields

$${}_5C_1(.04)^1(.96)^4 = .16987$$

Using combinations simplifies the determination of how many sequences are possible for a given value of x in a binomial distribution.

As another example, suppose 70% of all Americans believe cleaning up the environment is an important issue. What is the probability of randomly sampling four Americans and having exactly two of them say that they believe cleaning up the environment is an important issue? Let E represent the success of getting a person who believes cleaning up the environment is an important issue. For this example, $p = .70$. Let N represent the failure of not getting a person who believes cleaning up is an important issue (N denotes not important). The probability of getting one of these persons is $q = .30$.

The various sequences of getting two E's in a sample of four follow.

$$E_1, E_2, N_3, N_4$$
$$E_1, N_2, E_3, N_4$$
$$E_1, N_2, N_3, E_4$$
$$N_1, E_2, E_3, N_4$$
$$N_1, E_2, N_3, E_4$$
$$N_1, N_2, E_3, E_4$$

Two successes in a sample of four can occur six ways. Using combinations, the number of sequences is

$$_4C_2 = 6 \text{ ways}$$

The probability of selecting any individual sequence is

$$(.70)^2(.30)^2 = .0441$$

Thus the overall probability of getting exactly two people who believe cleaning up the environment is important out of four randomly selected people, when 70% of Americans believe cleaning up the environment is important, is

$$_4C_2(.70)^2(.30)^2 = .2646$$

Generalizing from these two examples yields the binomial formula, which can be used to solve binomial problems.

BINOMIAL FORMULA

$$P(x) = {}_nC_x \cdot p^x \cdot q^{n-x} = \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$$

where

$n =$ the number of trials (or the number being sampled)
$x =$ the number of successes desired
$p =$ the probability of getting a success in one trial
$q = 1 - p =$ the probability of getting a failure in one trial

The binomial formula summarizes the steps presented so far to solve binomial problems. The formula allows the solution of these problems quickly and efficiently.

DEMONSTRATION PROBLEM 5.2

A Gallup survey found that 65% of all financial consumers were very satisfied with their primary financial institution. Suppose that 25 financial consumers are sampled and if the Gallup survey result still holds true today, what is the probability that exactly 19 are very satisfied with their primary financial institution?

Solution

The value of p is .65 (very satisfied), the value of $q = 1 - p = 1 - .65 = .35$ (not very satisfied), $n = 25$, and $x = 19$. The binomial formula yields the final answer.

$$_{25}C_{19}(.65)^{19}(.35)^6 = (177{,}100)(.00027884)(.00183827) = .0908$$

If 65% of all financial consumers are very satisfied, about 9.08% of the time the researcher would get exactly 19 out of 25 financial consumers who are very satisfied with their financial institution. How many very satisfied consumers would one expect to get in 25 randomly selected financial consumers? If 65% of the financial consumers are very satisfied with their primary financial institution, one would expect to get about 65% of 25 or $(.65)(25) = 16.25$ very satisfied financial consumers. While in any individual sample of 25 the number of financial consumers who are very satisfied cannot be 16.25, business researchers understand the x values near 16.25 are the most likely occurrences.

DEMONSTRATION PROBLEM 5.3

According to the U.S. Census Bureau, approximately 6% of all workers in Jackson, Mississippi, are unemployed. In conducting a random telephone survey in Jackson, what is the probability of getting two or fewer unemployed workers in a sample of 20?

Solution

This problem must be worked as the union of three problems: (1) zero unemployed, $x = 0$; (2) one unemployed, $x = 1$; and (3) two unemployed, $x = 2$. In each problem, $p = .06$, $q = .94$, and $n = 20$. The binomial formula gives the following result.

$x = 0$		$x = 1$		$x = 2$	
${}_{20}C_0(.06)^0(.94)^{20}$	+	${}_{20}C_1(.06)^1(.94)^{19}$	+	${}_{20}C_2(.06)^2(.94)^{18}$	=
.2901	+	.3703	+	.2246	= .8850

If 6% of the workers in Jackson, Mississippi, are unemployed, the telephone surveyor would get zero, one, or two unemployed workers 88.5% of the time in a random sample of 20 workers. The requirement of getting two or fewer is satisfied by getting zero, one, or two unemployed workers. Thus this problem is the union of three probabilities. Whenever the binomial formula is used to solve for cumulative success (not an exact number), the probability of each x value must be solved and the probabilities summed. If an actual survey produced such a result, it would serve to validate the census figures.

Using the Binomial Table

Anyone who works enough binomial problems will begin to recognize that the probability of getting $x = 5$ successes from a sample size of $n = 18$ when $p = .10$ is the same no matter whether the five successes are left-handed people, defective parts, brand X purchasers, or any other variable. Whether the sample involves people, parts, or products does not matter in terms of the final probabilities. The essence of the problem is the same: $n = 18$, $x = 5$, and $p = .10$. Recognizing this fact, mathematicians constructed a set of binomial tables containing presolved probabilities.

Two parameters, n and p, describe or characterize a binomial distribution. Binomial distributions actually are a family of distributions. Every different value of n and/or every different value of p gives a different binomial distribution, and tables are available for various combinations of n and p values. Because of space limitations, the binomial tables presented in this text are limited. Table A.2 in Appendix A contains binomial tables. Each table is headed by a value of n. Nine values of p are presented in each table of size n. In the column below each value of p is the binomial distribution for that combination of n and p. Table 5.5 contains a segment of Table A.2 with the binomial probabilities for $n = 20$.

DEMONSTRATION PROBLEM 5.4

Solve the binomial probability for $n = 20$, $p = .40$, and $x = 10$ by using Table A.2, Appendix A.

Solution

To use Table A.2, first locate the value of n. Because $n = 20$ for this problem, the portion of the binomial tables containing values for $n = 20$ presented in Table 5.5 can be used. After locating the value of n, search horizontally across the top of the table for the appropriate value of p. In this problem, $p = .40$. The column under .40 contains the probabilities for the binomial distribution of $n = 20$ and $p = .40$. To get the probability of $x = 10$, find the value of x in the leftmost column and locate the probability in the table at the intersection of $p = .40$ and $x = 10$. The answer is .117. Working this problem by the binomial formula yields the same result.

$$ {}_{20}C_{10}(.40)^{10}(.60)^{10} = .1171 $$

TABLE 5.5

Excerpt from Table A.2, Appendix A

n = 20					Probability				
x	*.1*	*.2*	*.3*	*.4*	*.5*	*.6*	*.7*	*.8*	*.9*
0	.122	.012	.001	.000	.000	.000	.000	.000	.000
1	.270	.058	.007	.000	.000	.000	.000	.000	.000
2	.285	.137	.028	.003	.000	.000	.000	.000	.000
3	.190	.205	.072	.012	.001	.000	.000	.000	.000
4	.090	.218	.130	.035	.005	.000	.000	.000	.000
5	.032	.175	.179	.075	.015	.001	.000	.000	.000
6	.009	.109	.192	.124	.037	.005	.000	.000	.000
7	.002	.055	.164	.166	.074	.015	.001	.000	.000
8	.000	.022	.114	.180	.120	.035	.004	.000	.000
9	.000	.007	.065	.160	.160	.071	.012	.000	.000
10	.000	.002	.031	.117	.176	.117	.031	.002	.000
11	.000	.000	.012	.071	.160	.160	.065	.007	.000
12	.000	.000	.004	.035	.120	.180	.114	.022	.000
13	.000	.000	.001	.015	.074	.166	.164	.055	.002
14	.000	.000	.000	.005	.037	.124	.192	.109	.009
15	.000	.000	.000	.001	.015	.075	.179	.175	.032
16	.000	.000	.000	.000	.005	.035	.130	.218	.090
17	.000	.000	.000	.000	.001	.012	.072	.205	.190
18	.000	.000	.000	.000	.000	.003	.028	.137	.285
19	.000	.000	.000	.000	.000	.000	.007	.058	.270
20	.000	.000	.000	.000	.000	.000	.001	.012	.122

DEMONSTRATION PROBLEM 5.5

Decision Dilemma

According to Information Resources, which publishes data on market share for various products, Oreos control about 10% of the market for cookie brands. Suppose 20 purchasers of cookies are selected randomly from the population. What is the probability that fewer than four purchasers choose Oreos?

Solution

For this problem, $n = 20$, $p = .10$, and $x < 4$. Because $n = 20$, the portion of the binomial tables presented in Table 5.5 can be used to work this problem. Search along the row of p values for .10. Determining the probability of getting $x < 4$ involves summing the probabilities for $x = 0, 1, 2,$ and 3. The values appear in the x column at the intersection of each x value and $p = .10$.

x Value	Probability
0	.122
1	.270
2	.285
3	.190
	$(x < 4) = .867$

If 10% of all cookie purchasers prefer Oreos and 20 cookie purchasers are randomly selected, about 86.7% of the time fewer than four of the 20 will select Oreos.

Using the Computer to Produce a Binomial Distribution

Both Excel and Minitab can be used to produce the probabilities for virtually any binomial distribution. Such computer programs offer yet another option for solving binomial problems besides using the binomial formula or the binomial tables. Actually, the

TABLE 5.6

Minitab Output for the Binomial Distribution of $n = 23$, $p = .64$

PROBABILITY DENSITY FUNCTION

Binomial with n = 23 and p = 0.64

x	$P(X = x)$
0	0.000000
1	0.000000
2	0.000000
3	0.000001
4	0.000006
5	0.000037
6	0.000199
7	0.000858
8	0.003051
9	0.009040
10	0.022500
11	0.047273
12	0.084041
13	0.126420
14	0.160533
15	0.171236
16	0.152209
17	0.111421
18	0.066027
19	0.030890
20	0.010983
21	0.002789
22	0.000451
23	0.000035

computer packages in effect print out what would be a column of the binomial table. The advantages of using statistical software packages for this purpose are convenience (if the binomial tables are not readily available and a computer is) and the potential for generating tables for many more values than those printed in the binomial tables.

For example, a study of bank customers stated that 64% of all financial consumers believe banks are more competitive today than they were five years ago. Suppose 23 financial consumers are selected randomly and we want to determine the probabilities of various x values occurring. Table A.2 in Appendix A could not be used because only nine different p values are included and $p = .64$ is not one of those values. In addition, $n = 23$ is not included in the table. Without the computer, we are left with the binomial formula as the only option for solving binomial problems for $n = 23$ and $p = .64$. Particularly if the cumulative probability questions are asked (for example, $x \leq 10$), the binomial formula can be a tedious way to solve the problem.

Shown in Table 5.6 is the Minitab output for the binomial distribution of $n = 23$ and $p = .64$. With this computer output, a researcher could obtain or calculate the probability of any occurrence within the binomial distribution of $n = 23$ and $p = .64$. Table 5.7 contains Minitab output for the particular binomial problem, $P(x \leq 10)$ when $n = 23$ and $p = .64$, solved by using Minitab's cumulative probability capability.

Shown in Table 5.8 is Excel output for all values of x that have probabilities greater than .000001 for the binomial distribution discussed in Demonstration Problem 5.3 ($n = 20$, $p = .06$) and the solution to the question posed in Demonstration Problem 5.3.

Mean and Standard Deviation of a Binomial Distribution

A binomial distribution has an expected value or a long-run average, which is denoted by μ. The value of μ is determined by $n \cdot p$. For example, if $n = 10$ and $p = .4$, then $\mu = n \cdot p = (10)(.4) = 4$. The long-run average or expected value means that, if n items are sampled over and over for a long time and if p is the probability of getting a success on one trial, the average number of successes per sample is expected to be $n \cdot p$. If 40% of all graduate business students at a large university are women and if random samples of 10 graduate business students are selected many times, the expectation is that, on average, four of the 10 students would be women.

MEAN AND STANDARD DEVIATION OF A BINOMIAL DISTRIBUTION

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

Examining the mean of a binomial distribution gives an intuitive feeling about the likelihood of a given outcome.

According to one study, 64% of all financial consumers believe banks are more competitive today than they were five years ago. If 23 financial consumers are selected randomly, what is the expected number who believe banks are more competitive today than they were five years ago? This problem can be described by the binomial distribution of $n = 23$ and $p = .64$ given in Table 5.6. The mean of this binomial distribution yields the expected value for this problem.

TABLE 5.7

Minitab Output for the Binomial Problem, $P(x \leq 10 \mid n = 23 \text{ and } p = .64$

Cumulative Distribution Function

Binomial with n = 23 and p = 0.64

x	$P(X \Leftarrow x)$
10	0.0356916

TABLE 5.8

Excel Output for Demonstration Problem 5.3 and the Binomial Distribution of $n = 20$, $p = .06$

x	Prob(x)
0	0.2901
1	0.3703
2	0.2246
3	0.0860
4	0.0233
5	0.0048
6	0.0008
7	0.0001
8	0.0000
9	0.0000

The probability $x \le 2$ when $n = 20$ and $p = .06$ is .8850

$$\mu = n \cdot p = 23(.64) = 14.72$$

In the long run, if 23 financial consumers are selected randomly over and over and if indeed 64% of all financial consumers believe banks are more competitive today, then the experiment should average 14.72 financial consumers out of 23 who believe banks are more competitive today. Realize that because the binomial distribution is a discrete distribution you will never actually get 14.72 people out of 23 who believe banks are more competitive today. The mean of the distribution does reveal the relative likelihood of any individual occurrence. Examine Table 5.6. Notice that the highest probabilities are those near $x = 14.72$: $P(x = 15) = .1712$, $P(x = 14) = .1605$, and $P(x = 16) = .1522$. All other probabilities for this distribution are less than these probabilities.

The standard deviation of a binomial distribution is denoted σ and is equal to $\sqrt{n \cdot p \cdot q}$. The standard deviation for the financial consumer problem described by the binomial distribution in Table 5.6 is

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{(23)(.64)(.36)} = 2.30$$

Chapter 6 shows that some binomial distributions are nearly bell shaped and can be approximated by using the normal curve. The mean and standard deviation of a binomial distribution are the tools used to convert these binomial problems to normal curve problems.

Graphing Binomial Distributions

The graph of a binomial distribution can be constructed by using all the possible x values of a distribution and their associated probabilities. The x values usually are graphed along the x-axis and the probabilities are graphed along the y-axis.

Table 5.9 lists the probabilities for three different binomial distributions: $n = 8$ and $p = .20$, $n = 8$ and $p = .50$, and $n = 8$ and $p = .80$. Figure 5.2 displays Excel graphs for each of these three binomial distributions. Observe how the shape of the distribution changes as the value of p increases. For $p = .50$, the distribution is symmetrical. For $p = .20$ the distribution is skewed right and for $p = .80$ the distribution is skewed left. This pattern makes sense because the mean of the binomial distribution $n = 8$ and $p = .50$ is 4, which is in the middle of the distribution. The mean of the distribution $n = 8$ and $p = .20$ is 1.6, which results in the highest probabilities being near $x = 2$ and $x = 1$. This graph peaks early and stretches toward the higher values of x. The mean of the distribution $n = 8$ and $p = .80$ is 6.4, which results in the highest probabilities being near $x = 6$ and $x = 7$. Thus the peak of the distribution is nearer to 8 than to 0 and the distribution stretches back toward $x = 0$.

In any binomial distribution the largest x value that can occur is n and the smallest value is zero. Thus the graph of any binomial distribution is constrained by zero and n. If the p value of the distribution is not .50, this constraint will result in the graph "piling up" at one end and being skewed at the other end.

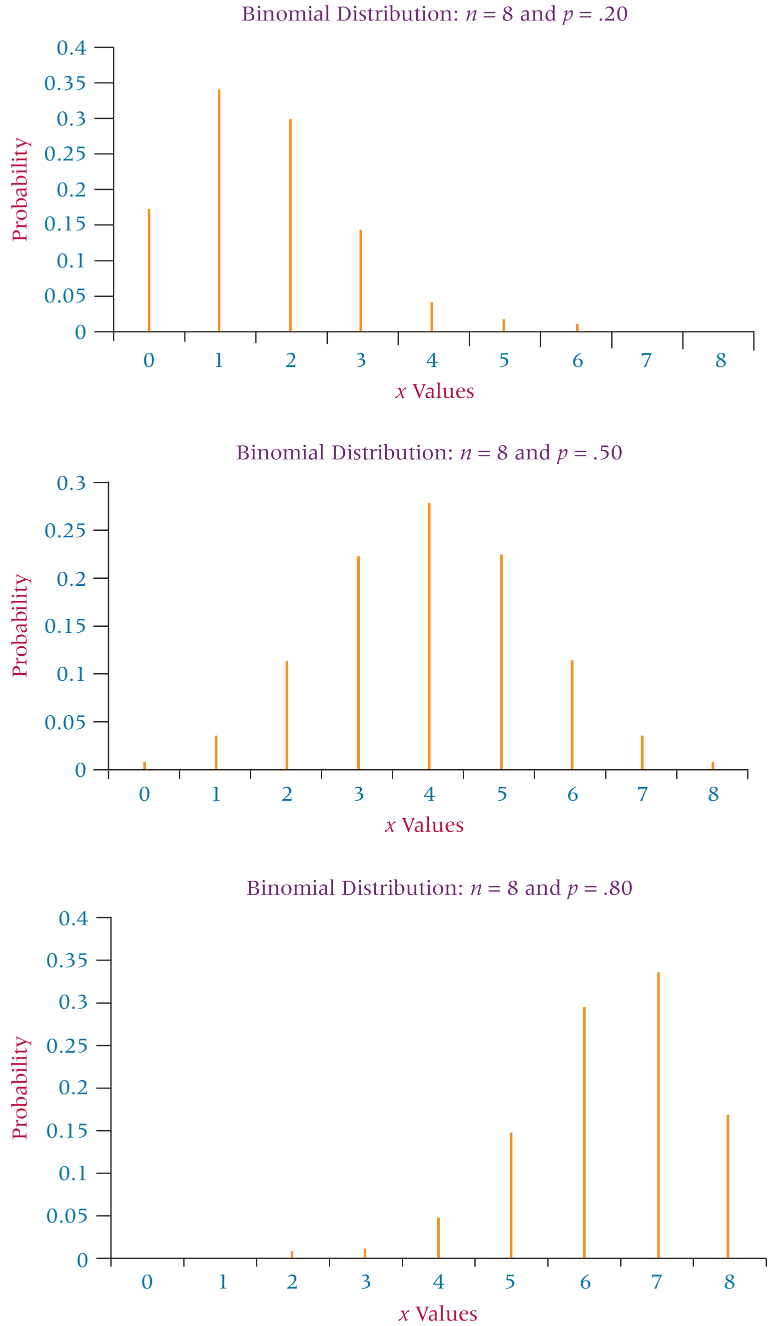
TABLE 5.9

Probabilities for Three Binomial Distributions with $n = 8$

	Probabilities for		
x	$p = .20$	$p = .50$	$p = .80$
0	.1678	.0039	.0000
1	.3355	.0312	.0001
2	.2936	.1094	.0011
3	.1468	.2187	.0092
4	.0459	.2734	.0459
5	.0092	.2187	.1468
6	.0011	.1094	.2936
7	.0001	.0312	.3355
8	.0000	.0039	.1678

FIGURE 5.2

Excel Graphs of Three Binomial Distributions with $n = 8$



DEMONSTRATION PROBLEM 5.6

A manufacturing company produces 10,000 plastic mugs per week. This company supplies mugs to another company, which packages the mugs as part of picnic sets. The second company randomly samples 10 mugs sent from the supplier. If two or fewer of the sampled mugs are defective, the second company accepts the lot. What is the probability that the lot will be accepted if the mug manufacturing company actually is producing mugs that are 10% defective? 20% defective? 30% defective? 40% defective?

Solution

In this series of binomial problems, $n = 10$, $x \leq 2$, and p ranges from .10 to .40. From Table A.2—and cumulating the values—we have the following probability of $x \leq 2$ for each p value and the expected value ($\mu = n \cdot p$).

p	Lot Accepted $P(x \leq 2)$	Expected Number of Defects (μ)
.10	.930	1.0
.20	.677	2.0
.30	.382	3.0
.40	.167	4.0

These values indicate that if the manufacturing company is producing 10% defective mugs, the probability is relatively high (.930) that the lot will be accepted by chance. For higher values of p, the probability of lot acceptance by chance decreases. In addition, as p increases, the expected value moves away from the acceptable values, $x \leq 2$. This move reduces the chances of lot acceptance.

STATISTICS IN BUSINESS TODAY

Plastic Bags vs. Bringing Your Own in Japan

In a move to protect and improve the environment, governments and companies around the world are making an effort to reduce the use of plastic bags by shoppers for transporting purchased food and goods. Specifically, in Yamagata City in northern Japan, the city concluded an agreement with seven local food supermarket chains to reduce plastic bag use in May of 2008 by having them agree to charge for the use of such bags. Before the agreement, in April of 2008, the average percentage of shoppers bringing their own shopping bags was about 35%. By the end of June, with some of the supermarket chains participating, the percentage had risen to almost 46%. However, by August, when 39 stores of the nine supermarket chains (two other chains joined the agreement) were charging for the use of plastic bags, the percentage rose to nearly 90%. It is estimated that the reduction of carbon dioxide emissions by this initiative is about 225 tons during July and August alone.

Source: http://www.japanfs.org/en/pages/028631.html

5.3 PROBLEMS

5.5 Solve the following problems by using the binomial formula.

a. If $n = 4$ and $p = .10$, find $P(x = 3)$.
b. If $n = 7$ and $p = .80$, find $P(x = 4)$.
c. If $n = 10$ and $p = .60$, find $P(x \geq 7)$.
d. If $n = 12$ and $p = .45$, find $P(5 \leq x \leq 7)$.

5.6 Solve the following problems by using the binomial tables (Table A.2).

a. If $n = 20$ and $p = .50$, find $P(x = 12)$.
b. If $n = 20$ and $p = .30$, find $P(x > 8)$.
c. If $n = 20$ and $p = .70$, find $P(x < 12)$.
d. If $n = 20$ and $p = .90$, find $P(x \leq 16)$.
e. If $n = 15$ and $p = .40$, find $P(4 \leq x \leq 9)$.
f. If $n = 10$ and $p = .60$, find $P(x \geq 7)$.

5.7 Solve for the mean and standard deviation of the following binomial distributions.

a. $n = 20$ and $p = .70$
b. $n = 70$ and $p = .35$
c. $n = 100$ and $p = .50$

5.8 Use the probability tables in Table A.2 and sketch the graph of each of the following binomial distributions. Note on the graph where the mean of the distribution falls.

a. $n = 6$ and $p = .70$

b. $n = 20$ and $p = .50$

c. $n = 8$ and $p = .80$

5.9 What is the first big change that American drivers made due to higher gas prices? According to an Access America survey, 30% said that it was cutting recreational driving. However, 27% said that it was consolidating or reducing errands. If these figures are true for all American drivers, and if 20 such drivers are randomly sampled and asked what is the first big change they made due to higher gas prices,

a. What is the probability that exactly 8 said that it was consolidating or reducing errands?

b. What is the probability that none of them said that it was cutting recreational driving?

c. What is the probability that more than 7 said that it was cutting recreational driving?

5.10 *The Wall Street Journal* reported some interesting statistics on the job market. One statistic is that 40% of all workers say they would change jobs for "slightly higher pay." In addition, 88% of companies say that there is a shortage of qualified job candidates. Suppose 16 workers are randomly selected and asked if they would change jobs for "slightly higher pay."

a. What is the probability that nine or more say yes?

b. What is the probability that three, four, five, or six say yes?

c. If 13 companies are contacted, what is the probability that exactly 10 say there is a shortage of qualified job candidates?

d. If 13 companies are contacted, what is the probability that all of the companies say there is a shortage of qualified job candidates?

e. If 13 companies are contacted, what is the expected number of companies that would say there is a shortage of qualified job candidates?

5.11 An increasing number of consumers believe they have to look out for themselves in the marketplace. According to a survey conducted by the Yankelovich Partners for *USA WEEKEND* magazine, 60% of all consumers have called an 800 or 900 telephone number for information about some product. Suppose a random sample of 25 consumers is contacted and interviewed about their buying habits.

a. What is the probability that 15 or more of these consumers have called an 800 or 900 telephone number for information about some product?

b. What is the probability that more than 20 of these consumers have called an 800 or 900 telephone number for information about some product?

c. What is the probability that fewer than 10 of these consumers have called an 800 or 900 telephone number for information about some product?

5.12 Studies have shown that about half of all workers who change jobs cash out their 401(k) plans rather than leaving the money in the account to grow. The percentage is much higher for workers with small 401(k) balances. In fact, 87% of workers with 401(k) accounts less than $5,000 opt to take their balance in cash rather than roll it over into individual retirement accounts when they change jobs.

a. Assuming that 50% of all workers who change jobs cash out their 401(k) plans, if 16 workers who have recently changed jobs that had 401(k) plans are randomly sampled, what is the probability that more than 10 of them cashed out their 401(k) plan?

b. If 10 workers who have recently changed jobs and had 401(k) plans with accounts less than \$5,000 are randomly sampled, what is the probability that exactly 6 of them cashed out?

5.13 In the past few years, outsourcing overseas has become more frequently used than ever before by U.S. companies. However, outsourcing is not without problems. A recent survey by *Purchasing* indicates that 20% of the companies that outsource overseas use a consultant. Suppose 15 companies that outsource overseas are randomly selected.

a. What is the probability that exactly five companies that outsource overseas use a consultant?

b. What is the probability that more than nine companies that outsource overseas use a consultant?

c. What is the probability that none of the companies that outsource overseas use a consultant?

d. What is the probability that between four and seven (inclusive) companies that outsource overseas use a consultant?

e. Construct a graph for this binomial distribution. In light of the graph and the expected value, explain why the probability results from parts (a) through (d) were obtained.

5.14 According to Cerulli Associates of Boston, 30% of all CPA financial advisors have an average client size between \$500,000 and \$1 million. Thirty-four percent have an average client size between \$1 million and \$5 million. Suppose a complete list of all CPA financial advisors is available and 18 are randomly selected from that list.

a. What is the expected number of CPA financial advisors that have an average client size between \$500,000 and \$1 million? What is the expected number with an average client size between \$1 million and \$5 million?

b. What is the probability that at least eight CPA financial advisors have an average client size between \$500,000 and \$1 million?

c. What is the probability that two, three, or four CPA financial advisors have an average client size between \$1 million and \$5 million?

d. What is the probability that none of the CPA financial advisors have an average client size between \$500,000 and \$1 million? What is the probability that none have an average client size between \$1 million and \$5 million? Which probability is higher and why?

5.4 POISSON DISTRIBUTION

Interactive Applet

The Poisson distribution is another discrete distribution. It is named after Simeon-Denis Poisson (1781–1840), a French mathematician, who published its essentials in a paper in 1837. The Poisson distribution and the binomial distribution have some similarities but also several differences. The binomial distribution describes a distribution of two possible outcomes designated as successes and failures from a given number of trials. The **Poisson distribution** *focuses only on the number of discrete occurrences over some interval or continuum.* A Poisson experiment does not have a given number of trials (n) as a binomial experiment does. For example, whereas a binomial experiment might be used to determine how many U.S.-made cars are in a random sample of 20 cars, a Poisson experiment might focus on the number of cars randomly arriving at an automobile repair facility during a 10-minute interval.

The Poisson distribution describes the occurrence of *rare events.* In fact, the Poisson formula has been referred to as the *law of improbable events.* For example, serious accidents at a chemical plant are rare, and the number per month might be described by the Poisson distribution. The Poisson distribution often is used to describe the number of random arrivals per some time interval. If the number of arrivals per interval is too frequent, the time interval can be reduced enough so that a rare number of occurrences is expected. Another example of a Poisson distribution is the number of random customer arrivals per five-minute interval at a small boutique on weekday mornings.

The Poisson distribution also has an application in the field of management science. The models used in queuing theory (theory of waiting lines) usually are based on the assumption that the Poisson distribution is the proper distribution to describe random arrival rates over a period of time.

The Poisson distribution has the following characteristics:

- It is a discrete distribution.
- It describes rare events.
- Each occurrence is independent of the other occurrences.
- It describes discrete occurrences over a continuum or interval.
- The occurrences in each interval can range from zero to infinity.
- The expected number of occurrences must hold constant throughout the experiment.

Examples of Poisson-type situations include the following:

1. Number of telephone calls per minute at a small business
2. Number of hazardous waste sites per county in the United States
3. Number of arrivals at a turnpike tollbooth per minute between 3 A.M. and 4 A.M. in January on the Kansas Turnpike
4. Number of sewing flaws per pair of jeans during production
5. Number of times a tire blows on a commercial airplane per week

Each of these examples represents a rare occurrence of events for some interval. Note that, although time is a more common interval for the Poisson distribution, intervals can range from a county in the United States to a pair of jeans. Some of the intervals in these examples might have zero occurrences. Moreover, the average occurrence per interval for many of these examples is probably in the single digits (1–9).

If a Poisson-distributed phenomenon is studied over a long period of time, a *long-run average* can be determined. This average is denoted **lambda (λ)**. Each Poisson problem contains a lambda value from which the probabilities of particular occurrences are determined. Although n and p are required to describe a binomial distribution, a Poisson distribution can be described by λ alone. The Poisson formula is used to compute the probability of occurrences over an interval for a given lambda value.

POISSON FORMULA

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where

$x = 0, 1, 2, 3, \ldots$
$\lambda =$ long-run average
$e = 2.718282$

Here, x is the number of occurrences per interval for which the probability is being computed, λ is the long-run average, and $e = 2.718282$ is the base of natural logarithms.

A word of caution about using the Poisson distribution to study various phenomena is necessary. The λ value must hold constant throughout a Poisson experiment. The researcher must be careful not to apply a given lambda to intervals for which lambda changes. For example, the average number of customers arriving at a Sears store during a one-minute interval will vary from hour to hour, day to day, and month to month. Different times of the day or week might produce different lambdas. The number of flaws per pair of jeans might vary from Monday to Friday. The researcher should be specific in describing the interval for which λ is being used.

Working Poisson Problems by Formula

Suppose bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of exactly 5 customers arriving in a 4-minute interval on a weekday afternoon? The lambda for this problem is 3.2 customers per 4 minutes. The value of x is 5 customers per 4 minutes. The probability of 5 customers randomly arriving during a 4-minute interval when the long-run average has been 3.2 customers per 4-minute interval is

$$\frac{(3.2^5)(e^{-3.2})}{5!} = \frac{(335.54)(.0408)}{120} = .1141$$

If a bank averages 3.2 customers every 4 minutes, the probability of 5 customers arriving during any one 4-minute interval is .1141.

DEMONSTRATION PROBLEM 5.7

Bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes. What is the probability of having more than 7 customers in a 4-minute interval on a weekday afternoon?

Solution

$$\lambda = 3.2 \text{ customers/minutes}$$
$$x > 7 \text{ customers/4 minutes}$$

In theory, the solution requires obtaining the values of $x = 8, 9, 10, 11, 12, 13, 14, \ldots \infty$. In actuality, each x value is determined until the values are so far away from $\lambda = 3.2$ that the probabilities approach zero. The exact probabilities are then summed to find $x > 7$.

$$P(x = 8|\lambda = 3.2) = \frac{(3.2^8)(e^{-3.2})}{8!} = .0111$$
$$P(x = 9|\lambda = 3.2) = \frac{(3.2^9)(e^{-3.2})}{9!} = .0040$$
$$P(x = 10|\lambda = 3.2) = \frac{(3.2^{10})(e^{-3.2})}{10!} = .0013$$
$$P(x = 11|\lambda = 3.2) = \frac{(3.2^{11})(e^{-3.2})}{11!} = .0004$$
$$P(x = 12|\lambda = 3.2) = \frac{(3.2^{12})(e^{-3.2})}{12!} = .0001$$
$$P(x = 13|\lambda = 3.2) = \frac{(3.2^{13})(e^{-3.2})}{13!} = .0000$$
$$P(x > 7) = P(x \geq 8) = .0169$$

If the bank has been averaging 3.2 customers every 4 minutes on weekday afternoons, it is unlikely that more than 7 people would randomly arrive in any one 4-minute period. This answer indicates that more than 7 people would randomly arrive in a 4-minute period only 1.69% of the time. Bank officers could use these results to help them make staffing decisions.

DEMONSTRATION PROBLEM 5.8

A bank has an average random arrival rate of 3.2 customers every 4 minutes. What is the probability of getting exactly 10 customers during an 8-minute interval?

Solution

$$\lambda = 3.2 \text{ customers}/4 \text{ minutes}$$
$$x = 10 \text{ customers}/8 \text{ minutes}$$

This example is different from the first two Poisson examples in that the intervals for lambda and the sample are different. The intervals must be the same in order to use λ and x together in the probability formula. The right way to approach this dilemma is to adjust the interval for lambda so that it and x have the same interval. The interval for x is 8 minutes, so lambda should be adjusted to an 8-minute interval. Logically, if the bank averages 3.2 customers every 4 minutes, it should average twice as many, or 6.4 customers, every 8 minutes. If x were for a 2-minute interval, the value of lambda would be halved from 3.2 to 1.6 customers per 2-minute interval. The wrong approach to this dilemma is to equalize the intervals by changing the x value. Never adjust or change x in a problem. Just because 10 customers arrive in one 8-minute interval does not mean that there would necessarily have been five customers in a 4-minute interval. There is no guarantee how the 10 customers are spread over the 8-minute interval. Always adjust the lambda value. After lambda has been adjusted for an 8-minute interval, the solution is

$$\lambda = 6.4 \text{ customers}/8 \text{ minutes}$$
$$x = 10 \text{ customers}/8 \text{ minutes}$$
$$\frac{(6.4)^{10}e^{-6.4}}{10!} = .0528$$

TABLE 5.10

Poisson Table for $\lambda = 1.6$

x	Probability
0	.2019
1	.3230
2	.2584
3	.1378
4	.0551
5	.0176
6	.0047
7	.0011
8	.0002
9	.0000

Using the Poisson Tables

Every value of lambda determines a different Poisson distribution. Regardless of the nature of the interval associated with a lambda, the Poisson distribution for a particular lambda is the same. Table A.3, Appendix A, contains the Poisson distributions for selected values of lambda. Probabilities are displayed in the table for each x value associated with a given lambda if the probability has a nonzero value to four decimal places. Table 5.10 presents a portion of Table A.3 that contains the probabilities of $x \leq 9$ if lambda is 1.6.

DEMONSTRATION PROBLEM 5.9

If a real estate office sells 1.6 houses on an average weekday and sales of houses on weekdays are Poisson distributed, what is the probability of selling exactly 4 houses in one day? What is the probability of selling no houses in one day? What is the probability of selling more than five houses in a day? What is the probability of selling 10 or more houses in a day? What is the probability of selling exactly 4 houses in two days?

Solution

$$\lambda = 1.6 \text{ houses/day}$$
$$P(x = 4|\lambda = 1.6) = ?$$

Table 5.10 gives the probabilities for $\lambda = 1.6$. The left column contains the x values. The line $x = 4$ yields the probability .0551. If a real estate firm has been averaging 1.6 houses sold per day, only 5.51% of the days would it sell exactly 4 houses and still maintain the lambda value. Line 1 of Table 5.10 shows the probability of selling no houses in a day (.2019). That is, on 20.19% of the days, the firm would sell no houses if sales are Poisson distributed with $\lambda = 1.6$ houses per day. Table 5.10 is not cumulative. To determine $P(x > 5)$, more than 5 houses, find the probabilities of $x = 6$, $x = 7$, $x = 8$, $x = 9$, . . . $x = ?$. However, at $x = 9$, the probability to four decimal places is zero, and Table 5.10 stops when an x value zeros out at four decimal places. The answer for $x > 5$ follows.

x	Probability
6	.0047
7	.0011
8	.0002
9	.0000
$x > 5 =$	.0060

What is the probability of selling 10 or more houses in one day? As the table zeros out at $x = 9$, the probability of $x \geq 10$ is essentially .0000—that is, if the real estate office has been averaging only 1.6 houses sold per day, it is virtually impossible to sell 10 or more houses in a day. What is the probability of selling exactly 4 houses in two days? In this case, the interval has been changed from one day to two days. Lambda is for one day, so an adjustment must be made: A lambda of 1.6 for one day converts to a lambda of 3.2 for two days. Table 5.10 no longer applies, so Table A.3 must be used to solve this problem. The answer is found by looking up $\lambda = 3.2$ and $x = 4$ in Table A.3: the probability is .1781.

Mean and Standard Deviation of a Poisson Distribution

The mean or expected value of a Poisson distribution is λ. It is the long-run average of occurrences for an interval if many random samples are taken. Lambda usually is not a whole number, so most of the time actually observing lambda occurrences in an interval is impossible.

For example, suppose $\lambda = 6.5$/interval for some Poisson-distributed phenomenon. The resulting numbers of x occurrences in 20 different random samples from a Poisson distribution with $\lambda = 6.5$ might be as follows.

6 9 7 4 8 7 6 6 10 6 5 5 8 4 5 8 5 4 9 10

Computing the mean number of occurrences from this group of 20 intervals gives 6.6. In theory, for infinite sampling the long-run average is 6.5. Note from the samples that, when λ is 6.5, several 5s and 6s occur. Rarely would sample occurrences of 1, 2, 3, 11, 12,

STATISTICS IN BUSINESS TODAY

Air Passengers' Complaints

In recent months, airline passengers have expressed much more dissatisfaction with airline service than ever before. Complaints include flight delays, lost baggage, long runway delays with little or no onboard service, overbooked flights, cramped space due to fuller flights, canceled flights, and grumpy airline employees. A majority of dissatisfied fliers merely grin and bear it. However, an increasing number of passengers log complaints with the U.S. Department of Transportation. In the mid-1990s, the average number of complaints per 100,000 passengers boarded was .66. In ensuing years, the average rose to .74, .86, 1.08, and 1.21.

In a recent year, according to the Department of Transportation, Southwest Airlines had the fewest average number of complaints per 100,000 with .27, followed by ExpressJet Airlines with .44, Alaska Airlines with .50, SkyWest Airlines with .53, and Frontier Airlines with .82. Within the top 10 largest U.S. airlines, U.S. Airways had the highest average number of complaints logged against it —2.11 complaints per 100,000 passengers.

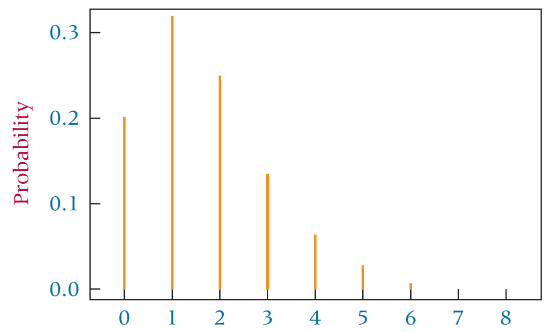
Because these average numbers are relatively small, it appears that the actual number of complaints per 100,000 is rare and may follow a Poisson distribution. In this case, λ represents the average number of complaints and the interval is 100,000 passengers. For example, using $\lambda = 1.21$ complaints (average for all airlines), if 100,000 boarded passengers were contacted, the probability that exactly three of them logged a complaint to the Department of Transportation could be computed as

$$\frac{(1.21)^3 e^{-1.21}}{3!} = .0880$$

That is, if 100,000 boarded passengers were contacted over and over, 8.80% of the time exactly three would have logged complaints with the Department of Transportation.

FIGURE 5.3

Minitab Graph of the Poisson Distribution for $\lambda = 1.6$



13, . . . occur when $\lambda = 6.5$. Understanding the mean of a Poisson distribution gives a feel for the actual occurrences that are likely to happen.

The variance of a Poisson distribution also is λ. The standard deviation is $\sqrt{\lambda}$. Combining the standard deviation with Chebyshev's theorem indicates the spread or dispersion of a Poisson distribution. For example, if $\lambda = 6.5$, the variance also is 6.5, and the standard deviation is 2.55. Chebyshev's theorem states that at least $1 - 1/k^2$ values are within k standard deviations of the mean. The interval $\mu \pm 2\sigma$ contains at least $1 - (1/2^2) = .75$ of the values. For $\mu = \lambda = 6.5$ and $\sigma = 2.55$, 75% of the values should be within the $6.5 \pm 2(2.55) = 6.5 \pm 5.1$ range. That is, the range from 1.4 to 11.6 should include at least 75% of all the values. An examination of the 20 values randomly generated for a Poisson distribution with $\lambda = 6.5$ shows that actually 100% of the values are within this range.

Graphing Poisson Distributions

The values in Table A.3, Appendix A, can be used to graph a Poisson distribution. The x values are on the x-axis and the probabilities are on the y-axis. Figure 5.3 is a Minitab graph for the distribution of values for $\lambda = 1.6$.

The graph reveals a Poisson distribution skewed to the right. With a mean of 1.6 and a possible range of x from zero to infinity, the values obviously will "pile up" at 0 and 1. Consider, however, the Minitab graph of the Poisson distribution for $\lambda = 6.5$ in Figure 5.4. Note that with $\lambda = 6.5$, the probabilities are greatest for the values of 5, 6, 7, and 8. The graph has less skewness, because the probability of occurrence of values near zero is small, as are the probabilities of large values of x.

Using the Computer to Generate Poisson Distributions

Using the Poisson formula to compute probabilities can be tedious when one is working problems with cumulative probabilities. The Poisson tables in Table A.3, Appendix A, are faster to use than the Poisson formula. However, Poisson tables are limited by the amount

FIGURE 5.4

Minitab Graph of the Poisson Distribution for $\lambda = 6.5$

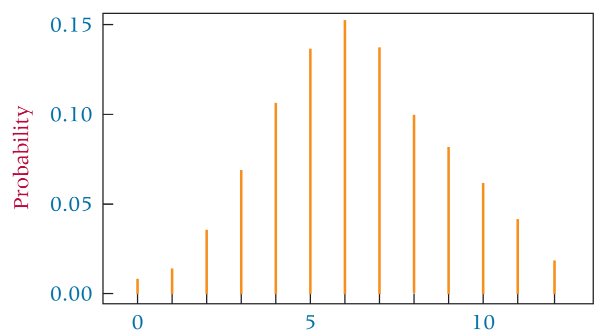


TABLE 5.11

Minitab Output for the Poisson Distribution $\lambda = 1.9$

```
PROBABILITY DENSITY FUNCTION

Poisson with mean = 1.9
 x          P(X = x)
 0          0.149569
 1          0.284180
 2          0.269971
 3          0.170982
 4          0.081216
 5          0.030862
 6          0.009773
 7          0.002653
 8          0.000630
 9          0.000133
10          0.000025
```

TABLE 5.12

Excel Output for the Poisson Distribution $\lambda = 1.6$

x	Probability
0	0.2019
1	0.3230
2	0.2584
3	0.1378
4	0.0551
5	0.0176
6	0.0047
7	0.0011
8	0.0002
9	0.0000

of space available, and Table A.3 only includes probability values for Poisson distributions with lambda values to the tenths place in most cases. For researchers who want to use lambda values with more precision or who feel that the computer is more convenient than textbook tables, some statistical computer software packages are an attractive option.

Minitab will produce a Poisson distribution for virtually any value of lambda. For example, one study by the National Center for Health Statistics claims that, on average, an American has 1.9 acute illnesses or injuries per year. If these cases are Poisson distributed, lambda is 1.9 per year. What does the Poisson probability distribution for this lambda look like? Table 5.11 contains the Minitab computer output for this distribution.

Excel can also generate probabilities of different values of x for any Poisson distribution. Table 5.12 displays the probabilities produced by Excel for the real estate problem from Demonstration Problem 5.9 using a lambda of 1.6.

Approximating Binomial Problems by the Poisson Distribution

Certain types of binomial distribution problems can be approximated by using the Poisson distribution. Binomial problems with large sample sizes and small values of p, which then generate rare events, are potential candidates for use of the Poisson distribution. As a rule of thumb, if $n > 20$ and $n \cdot p \leq 7$, the approximation is close enough to use the Poisson distribution for binomial problems.

If these conditions are met and the binomial problem is a candidate for this process, the procedure begins with computation of the mean of the binomial distribution, $\mu = n \cdot p$. Because μ is the expected value of the binomial, it translates to the expected value, λ, of the Poisson distribution. Using μ as the λ value and using the x value of the binomial problem allows approximation of the probability from a Poisson table or by the Poisson formula.

Large values of n and small values of p usually are not included in binomial distribution tables thereby precluding the use of binomial computational techniques. Using the Poisson distribution as an approximation to such a binomial problem in such cases is an attractive alternative; and indeed, when a computer is not available, it can be the only alternative.

As an example, the following binomial distribution problem can be worked by using the Poisson distribution: $n = 50$ and $p = .03$. What is the probability that $x = 4$? That is, $P(x = 4 \mid n = 50 \text{ and } p = .03) = ?$

To solve this equation, first determine lambda:

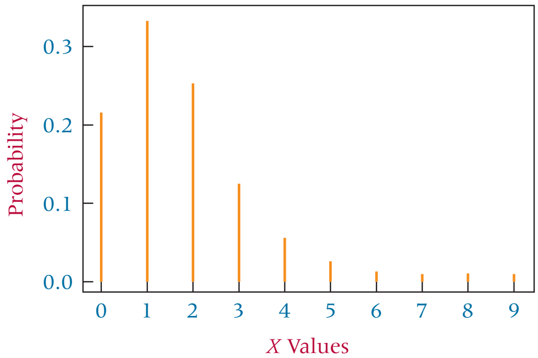
$$\lambda = \mu = n \cdot p = (50)(.03) = 1.5$$

As $n > 20$ and $n \cdot p \leq 7$, this problem is a candidate for the Poisson approximation. For $x = 4$, Table A.3 yields a probability of.0471 for the Poisson approximation. For comparison, working the problem by using the binomial formula yields the following results:

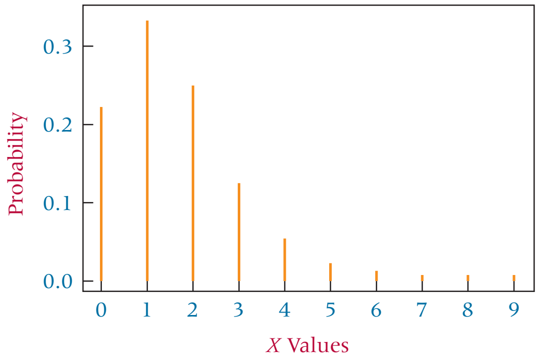
$$_{50}C_4(.03)^4(.97)^{46} = .0459$$

The Poisson approximation is .0012 different from the result obtained by using the binomial formula to work the problem.

A Minitab graph of this binomial distribution follows.



With $\lambda = 1.5$, the Poisson distribution can be generated. A Minitab graph of this Poisson distribution follows.



In comparing the two graphs, it is difficult to tell the difference between the binomial distribution and the Poisson distribution because the approximation of the binomial distribution by the Poisson distribution is close.

DEMONSTRATION PROBLEM 5.10

Suppose the probability of a bank making a mistake in processing a deposit is .0003. If 10,000 deposits (n) are audited, what is the probability that more than 6 mistakes were made in processing deposits?

Solution

$$\lambda = \mu = n \cdot p = (10{,}000)(.0003) = 3.0$$

Because $n > 20$ and $n \cdot p \leq 7$, the Poisson approximation is close enough to analyze $x > 6$. Table A.3 yields the following probabilities for $\lambda = 3.0$ and $x \geq 7$.

$\lambda = 3.0$

x	Probability
7	.0216
8	.0081
9	.0027
10	.0008
11	.0002
12	.0001
	$x > 6 = .0335$

To work this problem by using the binomial formula requires starting with $x = 7$.

$$_{10{,}000}C_7(.0003)^7(.9997)^{9993}$$

This process would continue for x values of 8, 9, 10, 11, . . . , until the probabilities approach zero. Obviously, this process is impractical, making the Poisson approximation an attractive alternative.

5.4 PROBLEMS

5.15 Find the following values by using the Poisson formula.

a. $P(x = 5 | \lambda = 2.3)$
b. $P(x = 2 | \lambda = 3.9)$
c. $P(x \le 3 | \lambda = 4.1)$
d. $P(x = 0 | \lambda = 2.7)$
e. $P(x = 1 | \lambda = 5.4)$
f. $P(4 < x < 8 | \lambda = 4.4)$

5.16 Find the following values by using the Poisson tables in Appendix A.

a. $P(x = 6 | \lambda = 3.8)$
b. $P(x > 7 | \lambda = 2.9)$
c. $P(3 \le x \le 9 | \lambda = 4.2)$
d. $P(x = 0 | \lambda = 1.9)$
e. $P(x \le 6 | \lambda = 2.9)$
f. $P(5 < x \le 8 | \lambda = 5.7)$

5.17 Sketch the graphs of the following Poisson distributions. Compute the mean and standard deviation for each distribution. Locate the mean on the graph. Note how the probabilities are graphed around the mean.

a. $\lambda = 6.3$
b. $\lambda = 1.3$
c. $\lambda = 8.9$
d. $\lambda = 0.6$

5.18 On Monday mornings, the First National Bank only has one teller window open for deposits and withdrawals. Experience has shown that the average number of arriving customers in a four-minute interval on Monday mornings is 2.8, and each teller can serve more than that number efficiently. These random arrivals at this bank on Monday mornings are Poisson distributed.

a. What is the probability that on a Monday morning exactly six customers will arrive in a four-minute interval?

b. What is the probability that no one will arrive at the bank to make a deposit or withdrawal during a four-minute interval?

c. Suppose the teller can serve no more than four customers in any four-minute interval at this window on a Monday morning. What is the probability that, during any given four-minute interval, the teller will be unable to meet the demand? What is the probability that the teller will be able to meet the demand? When demand cannot be met during any given interval, a second window is opened. What percentage of the time will a second window have to be opened?

d. What is the probability that exactly three people will arrive at the bank during a two-minute period on Monday mornings to make a deposit or a withdrawal? What is the probability that five or more customers will arrive during an eight-minute period?

5.19 A restaurant manager is interested in taking a more statistical approach to predicting customer load. She begins the process by gathering data. One of the restaurant hosts or hostesses is assigned to count customers every five minutes from 7 P.M. until 8 P.M. every Saturday night for three weeks. The data are shown here. After the data are gathered, the manager computes lambda using the data from all three weeks as one data set as a basis for probability analysis. What value of lambda did she find? Assume that these customers randomly arrive and that the arrivals are Poisson distributed. Use the value of lambda computed by the manager and help the manager calculate the probabilities in parts (a) through (e) for any given five-minute interval between 7 P.M. and 8 P.M. on Saturday night.

Number of Arrivals		
Week 1	**Week 2**	**Week 3**
3	1	5
6	2	3
4	4	5
6	0	3
2	2	5
3	6	4
1	5	7
5	4	3
1	2	4
0	5	8
3	3	1
3	4	3

a. What is the probability that no customers arrive during any given five-minute interval?

b. What is the probability that six or more customers arrive during any given five-minute interval?

c. What is the probability that during a 10-minute interval fewer than four customers arrive?

d. What is the probability that between three and six (inclusive) customers arrive in any 10-minute interval?

e. What is the probability that exactly eight customers arrive in any 15-minute interval?

5.20 According to the United National Environmental Program and World Health Organization, in Mumbai, India, air pollution standards for particulate matter are exceeded an average of 5.6 days in every three-week period. Assume that the distribution of number of days exceeding the standards per three-week period is Poisson distributed.

a. What is the probability that the standard is not exceeded on any day during a three-week period?

b. What is the probability that the standard is exceeded exactly six days of a three-week period?

c. What is the probability that the standard is exceeded 15 or more days during a three-week period? If this outcome actually occurred, what might you conclude?

5.21 The average number of annual trips per family to amusement parks in the United States is Poisson distributed, with a mean of 0.6 trips per year. What is the probability of randomly selecting an American family and finding the following?

a. The family did not make a trip to an amusement park last year.

b. The family took exactly one trip to an amusement park last year.

c. The family took two or more trips to amusement parks last year.

d. The family took three or fewer trips to amusement parks over a three-year period.

e. The family took exactly four trips to amusement parks during a six-year period.

5.22 Ship collisions in the Houston Ship Channel are rare. Suppose the number of collisions are Poisson distributed, with a mean of 1.2 collisions every four months.

a. What is the probability of having no collisions occur over a four-month period?

b. What is the probability of having exactly two collisions in a two-month period?

c. What is the probability of having one or fewer collisions in a six-month period? If this outcome occurred, what might you conclude about ship channel conditions during this period? What might you conclude about ship channel safety awareness during this period? What might you conclude about weather conditions during this period? What might you conclude about lambda?

5.23 A pen company averages 1.2 defective pens per carton produced (200 pens). The number of defects per carton is Poisson distributed.

a. What is the probability of selecting a carton and finding no defective pens?

b. What is the probability of finding eight or more defective pens in a carton?

c. Suppose a purchaser of these pens will quit buying from the company if a carton contains more than three defective pens. What is the probability that a carton contains more than three defective pens?

5.24 A medical researcher estimates that .00004 of the population has a rare blood disorder. If the researcher randomly selects 100,000 people from the population,

a. What is the probability that seven or more people will have the rare blood disorder?

b. What is the probability that more than 10 people will have the rare blood disorder?

c. Suppose the researcher gets more than 10 people who have the rare blood disorder in the sample of 100,000 but that the sample was taken from a particular geographic region. What might the researcher conclude from the results?

5.25 A data firm records a large amount of data. Historically, .9% of the pages of data recorded by the firm contain errors. If 200 pages of data are randomly selected,

a. What is the probability that six or more pages contain errors?

b. What is the probability that more than 10 pages contain errors?

c. What is the probability that none of the pages contain errors?

d. What is the probability that fewer than five pages contain errors?

5.26 A high percentage of people who fracture or dislocate a bone see a doctor for that condition. Suppose the percentage is 99%. Consider a sample in which 300 people are randomly selected who have fractured or dislocated a bone.

a. What is the probability that exactly five of them did not see a doctor?

b. What is the probability that fewer than four of them did not see a doctor?

c. What is the expected number of people who would not see a doctor?

5.5 HYPERGEOMETRIC DISTRIBUTION

Another discrete statistical distribution is the hypergeometric distribution. Statisticians often use the **hypergeometric distribution** to complement the types of analyses that can be made by using the binomial distribution. Recall that the binomial distribution applies, in theory, only to experiments in which the trials are done with replacement (independent events). The hypergeometric distribution applies only to experiments in which the trials are done without replacement.

The hypergeometric distribution, like the binomial distribution, consists of two possible outcomes: success and failure. However, the user must know the size of the population and the proportion of successes and failures in the population to apply the hypergeometric distribution. In other words, because the hypergeometric distribution is used when sampling is done without replacement, information about population makeup must be known in order to redetermine the probability of a success in each successive trial as the probability changes.

The hypergeometric distribution has the following characteristics:

- It is discrete distribution.
- Each outcome consists of either a success or a failure.

- Sampling is done without replacement.
- The population, N, is finite and known.
- The number of successes in the population, A, is known.

HYPERGEOMETRIC FORMULA

$$P(x) = \frac{{}_{A}C_{x} \cdot {}_{N-A}C_{n-x}}{{}_{N}C_{n}}$$

where

N = size of the population
n = sample size
A = number of successes in the population
x = number of successes in the sample; sampling is done *without* replacement

A hypergeometric distribution is characterized or described by three parameters: N, A, and n. Because of the multitude of possible combinations of these three parameters, creating tables for the hypergeometric distribution is practically impossible. Hence, the researcher who selects the hypergeometric distribution for analyzing data must use the hypergeometric formula to calculate each probability. Because this task can be tedious and time-consuming, most researchers use the hypergeometric distribution as a fallback position when working binomial problems without replacement. Even though the binomial distribution theoretically applies only when sampling is done with replacement and p stays constant, recall that, if the population is large enough in comparison with the sample size, the impact of sampling without replacement on p is minimal. Thus the binomial distribution can be used in some situations when sampling is done without replacement. Because of the tables available, using the binomial distribution instead of the hypergeometric distribution whenever possible is preferable. As a rule of thumb, if the sample size is less than 5% of the population, use of the binomial distribution rather than the hypergeometric distribution is acceptable when sampling is done without replacement. The hypergeometric distribution yields the exact probability, and the binomial distribution yields a good approximation of the probability in these situations.

In summary, the hypergeometric distribution should be used instead of the binomial distribution when the following conditions are present:

1. Sampling is being done without replacement.
2. $n \geq 5\% \, N$.

Hypergeometric probabilities are calculated under the assumption of equally likely sampling of the remaining elements of the sample space.

As an application of the hypergeometric distribution, consider the following problem. Twenty-four people, of whom eight are women, apply for a job. If five of the applicants are sampled randomly, what is the probability that exactly three of those sampled are women?

This problem contains a small, finite population of 24, or $N = 24$. A sample of five applicants is taken, or $n = 5$. The sampling is being done without replacement, because the five applicants selected for the sample are five different people. The sample size is 21% of the population, which is greater than 5% of the population ($n/N = 5/24 = .21$). The hypergeometric distribution is the appropriate distribution to use. The population breakdown is $A = 8$ women (successes) and $N - A = 24 - 8 = 16$ men. The probability of getting $x = 3$ women in the sample of $n = 5$ is

$$\frac{{}_{8}C_{3} \cdot {}_{16}C_{2}}{{}_{24}C_{5}} = \frac{(56)(120)}{42{,}504} = .1581$$

Conceptually, the combination in the denominator of the hypergeometric formula yields all the possible ways of getting n samples from a population, N, including the ones with the desired outcome. In this problem, there are 42,504 ways of selecting 5 people from 24 people. The numerator of the hypergeometric formula computes all the possible ways of getting x successes from the A successes available and $n - x$ failures from the $N - A$ available failures in the population. There are 56 ways of getting three women from a pool of eight, and there are 120 ways of getting 2 men from a pool of 16. The combinations of each are multiplied in the numerator because the joint probability of getting x successes and $n - x$ failures is being computed.

DEMONSTRATION PROBLEM 5.11

Suppose 18 major computer companies operate in the United States and that 12 are located in California's Silicon Valley. If three computer companies are selected randomly from the entire list, what is the probability that one or more of the selected companies are located in the Silicon Valley?

Solution

$$N = 18,\ n = 3,\ A = 12, \text{ and } x \geq 1$$

This problem is actually three problems in one: $x = 1$, $x = 2$, and $x = 3$. Sampling is being done without replacement, and the sample size is 16.6% of the population. Hence this problem is a candidate for the hypergeometric distribution. The solution follows.

$$\begin{array}{ccccccc} x = 1 & & x = 2 & & x = 3 & & \\ \frac{{}_{12}C_1 \cdot {}_6C_2}{{}_{18}C_3} & + & \frac{{}_{12}C_2 \cdot {}_6C_1}{{}_{18}C_3} & + & \frac{{}_{12}C_3 \cdot {}_6C_0}{{}_{18}C_3} & = & \\ .2206 & + & .4853 & + & .2696 & = & .9755 \end{array}$$

An alternative solution method using the law of complements would be one minus the probability that none of the companies is located in Silicon Valley, or

$$1 - P(x = 0 \mid N = 18,\ n = 3,\ A = 12)$$

Thus,

$$1 - \frac{{}_{12}C_0 \cdot {}_6C_3}{{}_{18}C_3} = 1 - .0245 = .9755$$

Using the Computer to Solve for Hypergeometric Distribution Probabilities

Using Minitab or Excel, it is possible to solve for hypergeometric distribution probabilities on the computer. Both software packages require the input of N, A, n, and x. In either package, the resulting output is the exact probability for that particular value of x. The Minitab output for the example presented in this section, where $N = 24$ people of whom $A = 8$ are women, $n = 5$ are randomly selected, and $x = 3$ are women, is displayed in Table 5.13. Note that Minitab represents successes in the population as "M." The Excel output for this same problem is presented in Table 5.14.

TABLE 5.13

Minitab Output for Hypergeometric Problem

```
PROBABILITY DENSITY FUNCTION

Hypergeometric with N = 24,
M = 8 and n = 24
x           P(X = x)
3           0.158103
```

TABLE 5.14

Excel Output for a Hypergeometric Problem

The probability of $x = 3$ when $N = 24$, $n = 5$, and $A = 8$ is: 0.158103

5.5 PROBLEMS

5.27 Compute the following probabilities by using the hypergeometric formula.

a. The probability of $x = 3$ if $N = 11$, $A = 8$, and $n = 4$

b. The probability of $x < 2$ if $N = 15$, $A = 5$, and $n = 6$

c. The probability of $x = 0$ if $N = 9$, $A = 2$, and $n = 3$

d. The probability of $x > 4$ if $N = 20$, $A = 5$, and $n = 7$

5.28 Shown here are the top 19 companies in the world in terms of oil refining capacity. Some of the companies are privately owned and others are state owned. Suppose six companies are randomly selected.

a. What is the probability that exactly one company is privately owned?

b. What is the probability that exactly four companies are privately owned?

c. What is the probability that all six companies are privately owned?

d. What is the probability that none of the companies is privately owned?

Company	Ownership Status
ExxonMobil	Private
Royal Dutch/Shell	Private
British Petroleum	Private
Sinopec	Private
Valero Energy	Private
Petroleos de Venezuela	State
Total	Private
ConocoPhillips	Private
China National	State
Saudi Arabian	State
Chevron	Private
Petroleo Brasilerio	State
Petroleos Mexicanos	State
National Iranian	State
OAO Yukos	Private
Nippon	Private
OAO Lukoil	Private
Repsol YPF	Private
Kuwait National	State

5.29 *Catalog Age* lists the top 17 U.S. firms in annual catalog sales. Dell Computer is number one followed by IBM and W. W. Grainger. Of the 17 firms on the list, 8 are in some type of computer-related business. Suppose four firms are randomly selected.

a. What is the probability that none of the firms is in some type of computer-related business?

b. What is the probability that all four firms are in some type of computer-related business?

c. What is the probability that exactly two are in non-computer-related business?

5.30 W. Edwards Deming in his red bead experiment had a box of 4,000 beads, of which 800 were red and 3,200 were white.* Suppose a researcher were to conduct a modified version of the red bead experiment. In her experiment, she has a bag of 20 beads, of which 4 are red and 16 are white. This experiment requires a participant to reach into the bag and randomly select five beads without replacement.

a. What is the probability that the participant will select exactly four white beads?

b. What is the probability that the participant will select exactly four red beads?

c. What is the probability that the participant will select all red beads?

*Mary Walton, "Deming's Parable of Red Beads," *Across the Board* (February 1987): 43–48.

5.31 Shown here are the top 10 U.S. cities ranked by number of rooms sold in a recent year.

Rank	City	Number of Rooms Sold
1	Las Vegas (NV)	40,000,000
2	Orlando (FL)	27,200,000
3	Los Angeles (CA)	25,500,000
4	Chicago (IL)	24,800,000
5	New York City (NY)	23,900,000
6	Washington (DC)	22,800,000
7	Atlanta (GA)	21,500,000
8	Dallas (TX)	15,900,000
9	Houston (TX)	14,500,000
10	San Diego (CA)	14,200,000

Suppose four of these cities are selected randomly.

a. What is the probability that exactly two cities are in California?

b. What is the probability that none of the cities is east of the Mississippi River?

c. What is the probability that exactly three of the cities are ones with more than 24 million rooms sold?

5.32 A company produces and ships 16 personal computers knowing that 4 of them have defective wiring. The company that purchased the computers is going to thoroughly test three of the computers. The purchasing company can detect the defective wiring. What is the probability that the purchasing company will find the following?

a. No defective computers

b. Exactly three defective computers

c. Two or more defective computers

d. One or fewer defective computer

5.33 A western city has 18 police officers eligible for promotion. Eleven of the 18 are Hispanic. Suppose only five of the police officers are chosen for promotion and that one is Hispanic. If the officers chosen for promotion had been selected by chance alone, what is the probability that one or fewer of the five promoted officers would have been Hispanic? What might this result indicate?

DECISION DILEMMA SOLVED

Life with a Cell Phone

Decision Dilemma Solved

Suppose that 14% of cell phone owners in the United States use only cellular phones. If 20 Americans are randomly selected, what is the probability that more than 7 use only cell phones? Converting the 14% to a proportion, the value of p is .14, and this is a classic binomial distribution problem with $n = 20$ and $x > 7$. Because the binomial distribution probability tables (Appendix A, Table A.2) do not include $p = .14$, the problem will have to be solved using the binomial formula for each of $x = 8, 9, 10, 11, \ldots, 20$.

$$\text{For } x = 8: \quad {}_{20}C_8(.14)^8(.86)^{12} = .0030$$

Solving for $x = 9$, 10, and 11 in a similar manner results in probabilities of .0007, .0001, and .0000, respectively. Since the probabilities "zero out" at $x = 11$, we need not proceed on to $x = 12, 13, 14, \ldots, 20$. Summing these four probabilities ($x = 8$, $x = 9$, $x = 10$, and $x = 11$) results in a total probability of .0038 as the answer to the posed question. To further understand these probabilities, we calculate the expected value of this distribution as:

$$\mu = n \cdot p = 20(.14) = 2.8$$

In the long run, one would expect to average about 2.8 Americans out of every 20 who consider their cell phone as their primary phone number. In light of this, there is a very small probability that more than seven Americans would do so.

The study also stated that 9 out of 10 cell users encounter others using their phones in an annoying way. Converting this to $p = .90$ and using $n = 25$ and $x < 20$, this, too, is a binomial problem, but it can be solved by using the binomial tables obtaining the values shown below:

x	Probability
19	.024
18	.007
17	.002
16	.000

The total of these probabilities is .033. Probabilities for all other values ($x \le 15$) are displayed as .000 in the binomial probability table and are not included here. If 90% of all cell phone users encounter others using their phones in an annoying way, the probability is very small (.033) that out of 25 randomly selected cell phone users less than 20 encounter others using their phones in an annoying way. The expected number in any random sample of 25 is $(25)(.90) = 22.5$.

Suppose, on average, cell phone users receive 3.6 calls per day. Given that information, what is the probability that a cell phone user receives no calls per day? Since random telephone calls are generally thought to be Poisson distributed, this problem can be solved by using either the Poisson probability formula or the Poisson tables (A.3, Appendix A). In this problem, $\lambda = 3.6$ and $x = 0$; and the probability associated with this is:

$$\frac{\lambda^x e^{-\lambda}}{x!} = \frac{(3.6)^0 e^{-3.6}}{0!} = .0273$$

What is the probability that a cell phone user receives 5 or more calls in a day? Since this is a cumulative probability question ($x \ge 5$), the best option is to use the Poisson probability tables (A.3, Appendix A) to obtain:

x	Probability
5	.1377
6	.0826
7	.0425
8	.0191
9	.0076
10	.0028
11	.0009
12	.0003
13	.0001
14	.0000
total	.2936

There is a 29.36% chance that a cell phone user will receive 5 or more calls per day if, on average, such a cell phone user averages 3.6 calls per day.

SUMMARY

Probability experiments produce random outcomes. A variable that contains the outcomes of a random experiment is called a random variable. Random variables such that the set of all possible values is at most a finite or countably infinite number of possible values are called discrete random variables. Random variables that take on values at all points over a given interval are called continuous random variables. Discrete distributions are constructed from discrete random variables. Continuous distributions are constructed from continuous random variables. Three discrete distributions are the binomial distribution, Poisson distribution, and hypergeometric distribution.

The binomial distribution fits experiments when only two mutually exclusive outcomes are possible. In theory, each trial in a binomial experiment must be independent of the other trials. However, if the population size is large enough in relation to the sample size ($n < 5\%N$), the binomial distribution can be used where applicable in cases where the trials are not independent. The probability of getting a desired outcome on any one trial is denoted as p, which is the probability of getting a success. The binomial formula is used to determine the probability of obtaining x outcomes in n trials. Binomial distribution problems can be solved more rapidly with the use of binomial tables than by formula. Table A.2 of Appendix A contains binomial tables for selected values of n and p.

The Poisson distribution usually is used to analyze phenomena that produce rare occurrences. The only information required to generate a Poisson distribution is the long-run average, which is denoted by lambda (λ). The Poisson distribution pertains to occurrences over some interval. The assumptions are that each occurrence is independent of other occurrences and that the value of lambda remains constant throughout the experiment. Poisson probabilities can be determined by either the Poisson formula or the Poisson tables in Table A.3 of Appendix A. The Poisson distribution can be used to approximate binomial distribution problems when n is large ($n > 20$), p is small, and $n \cdot p \le 7$.

The hypergeometric distribution is a discrete distribution that is usually used for binomial-type experiments when the population is small and finite and sampling is done without replacement. Because using the hypergeometric distribution is a tedious process, using the binomial distribution whenever possible is generally more advantageous.

KEY TERMS

Flash Cards

binomial distribution
continuous distributions
continuous random variables
discrete distributions
discrete random variables
hypergeometric distribution
lambda (λ)
mean or expected value
Poisson distribution
random variable

FORMULAS

Mean (expected) value of a discrete distribution

$$\mu = E(x) = \Sigma[x \cdot P(x)]$$

Variance of a discrete distribution

$$\sigma^2 = \Sigma[(x - \mu)^2 \cdot P(x)]$$

Standard deviation of a discrete distribution

$$\sigma = \sqrt{\Sigma[(x - \mu)^2 \cdot P(x)]}$$

Binomial formula

$$_nC_x \cdot p^x \cdot q^{n-x} = \frac{n!}{x!(n-x)!} \cdot p^x \cdot q^{n-x}$$

Mean of a binomial distribution

$$\mu = n \cdot p$$

Standard deviation of a binomial distribution

$$\sigma = \sqrt{n \cdot p \cdot q}$$

Poisson Formula

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Hypergeometric formula

$$P(x) = \frac{_AC_x \cdot {}_{N-A}C_{n-x}}{_NC_n}$$

ETHICAL CONSIDERATIONS

Several points must be emphasized about the use of discrete distributions to analyze data. The independence and/or size assumptions must be met in using the binomial distribution in situations where sampling is done without replacement. Size and λ assumptions must be satisfied in using the Poisson distribution to approximate binomial problems. In either case, failure to meet such assumptions can result in spurious conclusions.

As n increases, the use of binomial distributions to study exact x-value probabilities becomes questionable in decision making. Although the probabilities are mathematically correct, as n becomes larger, the probability of any particular x value becomes lower because there are more values among which to split the probabilities. For example, if $n = 100$ and $p = .50$, the probability of $x = 50$ is .0796. This probability of occurrence appears quite low, even though $x = 50$ is the expected value of this distribution and is also the value most likely to occur. It is more useful to decision makers and, in a sense, probably more ethical to present cumulative values for larger sizes of n. In this example, it is probably more useful to examine $P(x > 50)$ than $P(x = 50)$.

The reader is warned in the chapter that the value of λ is assumed to be constant in a Poisson distribution experiment. Researchers may produce spurious results because the λ value changes during a study. For example, suppose the value of λ is obtained for the number of customer arrivals at a toy store between 7 P.M. and 9 P.M. in the month of December. Because December is an active month in terms of traffic volume through a toy store, the use of such a λ to analyze arrivals at the same store between noon and 2 P.M. in February would be inappropriate and, in a sense, unethical.

Errors in judgment such as these are probably more a case of misuse than lack of ethics. However, it is important that statisticians and researchers adhere to assumptions and appropriate applications of these techniques. The inability or unwillingness to do so opens the way for unethical decision making.

SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

5.34 Solve for the probabilities of the following binomial distribution problems by using the binomial formula.

a. If $n = 11$ and $p = .23$, what is the probability that $x = 4$?
b. If $n = 6$ and $p = .50$, what is the probability that $x \geq 1$?
c. If $n = 9$ and $p = .85$, what is the probability that $x > 7$?
d. If $n = 14$ and $p = .70$, what is the probability that $x \leq 3$?

5.35 Use Table A.2, Appendix A, to find the values of the following binomial distribution problems.

a. $P(x = 14 \mid n = 20 \text{ and } p = .60)$
b. $P(x < 5 \mid n = 10 \text{ and } p = .30)$
c. $P(x \geq 12 \mid n = 15 \text{ and } p = .60)$
d. $P(x > 20 \mid n = 25 \text{ and } p = .40)$

5.36 Use the Poisson formula to solve for the probabilities of the following Poisson distribution problems.

a. If $\lambda = 1.25$, what is the probability that $x = 4$?

b. If $\lambda = 6.37$, what is the probability that $x \leq 1$?

c. If $\lambda = 2.4$, what is the probability that $x > 5$?

5.37 Use Table A.3, Appendix A, to find the following Poisson distribution values.

a. $P(x = 3 | \lambda = 1.8)$

b. $P(x < 5 | \lambda = 3.3)$

c. $P(x \geq 3 | \lambda = 2.1)$

d. $P(2 < x \leq 5 | \lambda = 4.2)$

5.38 Solve the following problems by using the hypergeometric formula.

a. If $N = 6$, $n = 4$, and $A = 5$, what is the probability that $x = 3$?

b. If $N = 10$, $n = 3$, and $A = 5$, what is the probability that $x \leq 1$?

c. If $N = 13$, $n = 5$, and $A = 3$, what is the probability that $x \geq 2$?

TESTING YOUR UNDERSTANDING

5.39 In a study by Peter D. Hart Research Associates for the Nasdaq Stock Market, it was determined that 20% of all stock investors are retired people. In addition, 40% of all U.S. adults invest in mutual funds. Suppose a random sample of 25 stock investors is taken.

a. What is the probability that exactly seven are retired people?

b. What is the probability that 10 or more are retired people?

c. How many retired people would you expect to find in a random sample of 25 stock investors?

d. Suppose a random sample of 20 U.S. adults is taken. What is the probability that exactly eight adults invested in mutual funds?

e. Suppose a random sample of 20 U.S. adults is taken. What is the probability that fewer than six adults invested in mutual funds?

f. Suppose a random sample of 20 U.S. adults is taken. What is the probability that none of the adults invested in mutual funds?

g. Suppose a random sample of 20 U.S. adults is taken. What is the probability that 12 or more adults invested in mutual funds?

h. For parts e–g, what exact number of adults would produce the highest probability? How does this compare to the expected number?

5.40 A service station has a pump that distributes diesel fuel to automobiles. The station owner estimates that only about 3.2 cars use the diesel pump every 2 hours. Assume the arrivals of diesel pump users are Poisson distributed.

a. What is the probability that three cars will arrive to use the diesel pump during a one-hour period?

b. Suppose the owner needs to shut down the diesel pump for half an hour to make repairs. However, the owner hates to lose any business. What is the probability that no cars will arrive to use the diesel pump during a half-hour period?

c. Suppose five cars arrive during a one-hour period to use the diesel pump. What is the probability of five or more cars arriving during a one-hour period to use the diesel pump? If this outcome actually occurred, what might you conclude?

5.41 In a particular manufacturing plant, two machines (A and B) produce a particular part. One machine (B) is newer and faster. In one five-minute period, a lot consisting of 32 parts is produced. Twenty-two are produced by machine B and the rest by machine A. Suppose an inspector randomly samples a dozen of the parts from this lot.

a. What is the probability that exactly three parts were produced by machine A?

b. What is the probability that half of the parts were produced by each machine?

c. What is the probability that all of the parts were produced by machine B?

d. What is the probability that seven, eight, or nine parts were produced by machine B?

5.42 Suppose that, for every lot of 100 computer chips a company produces, an average of 1.4 are defective. Another company buys many lots of these chips at a time, from which one lot is selected randomly and tested for defects. If the tested lot contains more than three defects, the buyer will reject all the lots sent in that batch. What is the probability that the buyer will accept the lots? Assume that the defects per lot are Poisson distributed.

5.43 The National Center for Health Statistics reports that 25% of all Americans between the ages of 65 and 74 have a chronic heart condition. Suppose you live in a state where the environment is conducive to good health and low stress and you believe the conditions in your state promote healthy hearts. To investigate this theory, you conduct a random telephone survey of 20 persons 65 to 74 years of age in your state.

a. On the basis of the figure from the National Center for Health Statistics, what is the expected number of persons 65 to 74 years of age in your survey who have a chronic heart condition?

b. Suppose only one person in your survey has a chronic heart condition. What is the probability of getting one or fewer people with a chronic heart condition in a sample of 20 if 25% of the population in this age bracket has this health problem? What do you conclude about your state from the sample data?

5.44 A survey conducted for the Northwestern National Life Insurance Company revealed that 70% of American workers say job stress caused frequent health problems. One in three said they expected to burn out in the job in the near future. Thirty-four percent said they thought seriously about quitting their job last year because of work-place stress. Fifty-three percent said they were required to work more than 40 hours a week very often or somewhat often.

a. Suppose a random sample of 10 American workers is selected. What is the probability that more than seven of them say job stress caused frequent health problems? What is the expected number of workers who say job stress caused frequent health problems?

b. Suppose a random sample of 15 American workers is selected. What is the expected number of these sampled workers who say they will burn out in the near future? What is the probability that none of the workers say they will burn out in the near future?

c. Suppose a sample of seven workers is selected randomly. What is the probability that all seven say they are asked very often or somewhat often to work more than 40 hours a week? If this outcome actually happened, what might you conclude?

5.45 According to Padgett Business Services, 20% of all small-business owners say the most important advice for starting a business is to prepare for long hours and hard work. Twenty-five percent say the most important advice is to have good financing ready. Nineteen percent say having a good plan is the most important advice; 18% say studying the industry is the most important advice; and 18% list other advice. Suppose 12 small business owners are contacted, and assume that the percentages hold for all small-business owners.

a. What is the probability that none of the owners would say preparing for long hours and hard work is the most important advice?

b. What is the probability that six or more owners would say preparing for long hours and hard work is the most important advice?

c. What is the probability that exactly five owners would say having good financing ready is the most important advice?

d. What is the expected number of owners who would say having a good plan is the most important advice?

5.46 According to a recent survey, the probability that a passenger files a complaint with the Department of Transportation about a particular U.S. airline is .000014. Suppose 100,000 passengers who flew with this particular airline are randomly contacted.

a. What is the probability that exactly five passengers filed complaints?

b. What is the probability that none of the passengers filed complaints?

c. What is the probability that more than six passengers filed complaints?

5.47 A hair stylist has been in business one year. Sixty percent of his customers are walk-in business. If he randomly samples eight of the people from last week's list of customers, what is the probability that three or fewer were walk-ins? If this outcome actually occurred, what would be some of the explanations for it?

5.48 A Department of Transportation survey showed that 60% of U.S. residents over 65 years of age oppose use of cell phones in flight even if there were no issues with the phones interfering with aircraft communications systems. If this information is correct and if a researcher randomly selects 25 U.S. residents who are over 65 years of age,

a. What is the probability that exactly 12 oppose the use of cell phones in flight?

b. What is the probability that more than 17 oppose the use of cell phones in flight?

c. What is the probability that less than eight oppose the use of cell phones in flight? If the researcher actually got less than eight, what might she conclude about the Department of Transportation survey?

5.49 A survey conducted by the Consumer Reports National Research Center reported, among other things, that women spend an average of 1.2 hours per week shopping online. Assume that hours per week shopping online are Poisson distributed. If this survey result is true for all women and if a woman is randomly selected,

a. What is the probability that she did not shop at all online over a one-week period?

b. What is the probability that a woman would shop three or more hours online during a one-week period?

c. What is the probability that a woman would shop fewer than five hours in a three-week period?

5.50 According to the Audit Bureau of Circulations, the top 25 city newspapers in the United States ranked according to circulation are:

Rank	Newspaper
1	New York Times (NY)
2	Los Angeles Times (CA)
3	New York Daily News (NY)
4	New York Post (NY)
5	Washington Post (DC)
6	Chicago Tribune (IL)
7	Houston Chronicle (TX)
8	Phoenix Arizona Republic (AZ)
9	Long Island Newsday (NY)
10	San Francisco Chronicle (CA)
11	Dallas Morning News (TX)
12	Boston Globe (MA)
13	Newark Star-Ledger (NJ)
14	Philadelphia Inquirer (PA)
15	Cleveland Plain Dealer (OH)
16	Atlanta Journal-Constitution (GA)
17	Minneapolis Star Tribune (MN)
18	St. Petersburg Times (FL)
19	Chicago Sun-Times (IL)
20	Detroit Free Press (MI)
21	Portland Oregonian (OR)
22	San Diego Union-Tribune (CA)
23	Sacramento Bee (CA)
24	Indianapolis Star (IN)
25	St. Louis Post Dispatch (MO)

Suppose a researcher wants to sample a portion of these newspapers and compare the sizes of the business sections of the Sunday papers. She randomly samples eight of these newspapers.

a. What is the probability that the sample contains exactly one newspaper located in New York state?

b. What is the probability that half of the newspapers are ranked in the top 10 by circulation?

c. What is the probability that none of the newspapers is located in California?

d. What is the probability that exactly three of the newspapers are located in states that begin with the letter *M*?

5.51 An office in Albuquerque has 24 workers including management. Eight of the workers commute to work from the west side of the Rio Grande River. Suppose six of the office workers are randomly selected.

a. What is the probability that all six workers commute from the west side of the Rio Grande?

b. What is the probability that none of the workers commute from the west side of the Rio Grande?

c. Which probability from parts (a) and (b) was greatest? Why do you think this is?

d. What is the probability that half of the workers do not commute from the west side of the Rio Grande?

5.52 According to the U.S. Census Bureau, 20% of the workers in Atlanta use public transportation. If 25 Atlanta workers are randomly selected, what is the expected number to use public transportation? Graph the binomial distribution for this sample. What are the mean and the standard deviation for this distribution? What is the probability that more than 12 of the selected workers use public transportation? Explain conceptually and from the graph why you would get this probability. Suppose you randomly sample 25 Atlanta workers and actually get 14 who use public transportation. Is this outcome likely? How might you explain this result?

5.53 One of the earliest applications of the Poisson distribution was in analyzing incoming calls to a telephone switchboard. Analysts generally believe that random phone calls are Poisson distributed. Suppose phone calls to a switchboard arrive at an average rate of 2.4 calls per minute.

a. If an operator wants to take a one-minute break, what is the probability that there will be no calls during a one-minute interval?

b. If an operator can handle at most five calls per minute, what is the probability that the operator will be unable to handle the calls in any one-minute period?

c. What is the probability that exactly three calls will arrive in a two-minute interval?

d. What is the probability that one or fewer calls will arrive in a 15-second interval?

5.54 A survey by Frank N. Magid Associates revealed that 3% of Americans are not connected to the Internet at home. Another researcher randomly selects 70 Americans.

a. What is the expected number of these who would not be connected to the Internet at home?

b. What is the probability that eight or more are not connected to the Internet at home?

c. What is the probability that between three and six (inclusive) are not connected to the Internet at home?

5.55 Suppose that in the bookkeeping operation of a large corporation the probability of a recording error on any one billing is .005. Suppose the probability of a recording error from one billing to the next is constant, and 1,000 billings are randomly sampled by an auditor.

a. What is the probability that fewer than four billings contain a recording error?

b. What is the probability that more than 10 billings contain a billing error?

c. What is the probability that all 1,000 billings contain no recording errors?

5.56 According to the American Medical Association, about 36% of all U.S. physicians under the age of 35 are women. Your company has just hired eight physicians under the age of 35 and none is a woman. If a group of women physicians under the age of 35 want to sue your company for discriminatory hiring practices, would they have a strong case based on these numbers? Use the binomial distribution to determine the probability of the company's hiring result occurring randomly, and comment on the potential justification for a lawsuit.

5.57 The following table lists the 25 largest U.S. universities according to enrollment figures from *The World Almanac*. The state of location is given in parentheses.

University	Enrollment
University of Phoenix (AZ)	160,150
Ohio State University (OH)	51,818
Arizona State University (AZ)	51,234
University of Florida (FL)	50,822
University of Minnesota (MN)	50,402
University of Texas at Austin (TX)	49,697
University of Central Florida (FL)	46,719
Michigan State University (MI)	45,520
Texas A&M University (TX)	45,380
University of South Florida (FL)	43,636
Penn State University Park (PA)	42,914
University of Illinois (IL)	42,728
University of Wisconsin (WI)	41,466
New York University (NY)	40,870
University of Michigan (MI)	40,025
Florida State University (FL)	39,973
University of Washington (WA)	39,524
Purdue University (IN)	39,228
Indiana University (IN)	38,247
University of California, Los Angeles (CA)	38,218
Florida International University (FL)	37,997
University of Arizona (AZ)	36,805
California State University, Fullerton (CA)	35,921
California State University, Long Beach (CA)	35,574
University of Maryland, College Park (MD)	35,300

a. If five different universities are selected randomly from the list, what is the probability that three of them have enrollments of 40,000 or more?
b. If eight different universities are selected randomly from the list, what is the probability that two or fewer are universities in Florida?
c. Suppose universities are being selected randomly from this list with replacement. If five universities are sampled, what is the probability that the sample will contain exactly two universities in California?

5.58 In one midwestern city, the government has 14 repossessed houses, which are evaluated to be worth about the same. Ten of the houses are on the north side of town and the rest are on the west side. A local contractor submitted a bid to purchase four of the houses. Which houses the contractor will get is subject to a random draw.
a. What is the probability that all four houses selected for the contractor will be on the north side of town?
b. What is the probability that all four houses selected for the contractor will be on the west side of town?
c. What is the probability that half of the houses selected for the contractor will be on the west side and half on the north side of town?

5.59 The Public Citizen's Health Research Group studied the serious disciplinary actions that were taken during a recent year on nonfederal medical doctors in the United States. The national average was 2.92 serious actions per 1,000 doctors. The state with the lowest number was South Carolina, with 1.18 serious actions per 1,000 doctors. Assume that the numbers of serious actions per 1,000 doctors in both the United States and in South Carolina are Poisson distributed.
a. What is the probability of randomly selecting 1,000 U.S. doctors and finding no serious actions taken?
b. What is the probability of randomly selecting 2,000 U.S. doctors and finding six serious actions taken?
c. What is the probability of randomly selecting 3,000 South Carolina doctors and finding fewer than seven serious actions taken?

INTERPRETING THE OUTPUT

5.60 Study the Minitab output. Discuss the type of distribution, the mean, standard deviation, and why the probabilities fall as they do.

Probability Density Function

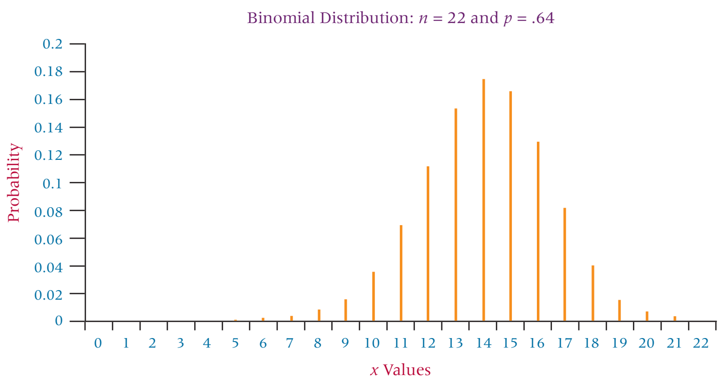
Binomial with n = 15 and n = 0.36

x	$P(X = x)$
0	0.001238
1	0.010445
2	0.041128
3	0.100249
4	0.169170
5	0.209347
6	0.196263
7	0.141940
8	0.079841
9	0.034931
10	0.011789
11	0.003014
12	0.000565
13	0.000073
14	0.000006
15	0.000000

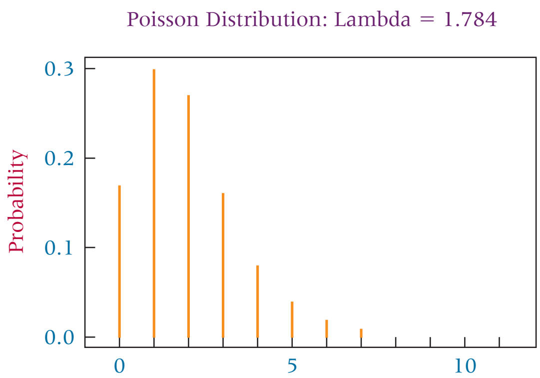
5.61 Study the Excel output. Explain the distribution in terms of shape and mean. Are these probabilities what you would expect? Why or why not?

x Values	Poisson Probabilities: $\lambda = 2.78$
0	0.0620
1	0.1725
2	0.2397
3	0.2221
4	0.1544
5	0.0858
6	0.0398
7	0.0158
8	0.0055
9	0.0017
10	0.0005
11	0.0001

5.62 Study the graphical output from Excel. Describe the distribution and explain why the graph takes the shape it does.



5.63 Study the Minitab graph. Discuss the distribution including type, shape, and probability outcomes.



see www.wiley.com/college/black

ANALYZING THE DATABASES

1. Use the Consumer Food database. What proportion of the database households are in the Metro area? Use this as the value of p in a binomial distribution. If you were to randomly select 12 of these households, what is the probability that fewer than 3 would be households in the Metro area? If you were to randomly select 25 of these households, what is the probability that exactly 8 would be in the Metro area?
2. Use the Hospital database. What is the breakdown between hospitals that are general medical hospitals and those that are psychiatric hospitals in this database of 200 hospitals? (Hint: In Service, 1 = general medical and 2 = psychiatric). Using these figures and the hypergeometric distribution, determine the probability of randomly selecting 16 hospitals from the database and getting exactly 9 that are psychiatric hospitals. Now, determine the number of hospitals in this database that are for-profit (Hint: In Control, 3 = for-profit). From this number, calculate p, the proportion of hospitals that are for-profit. Using this value of p and the binomial distribution, determine the probability of randomly selecting 30 hospitals and getting exactly 10 that are for-profit.

CASE

KODAK TRANSITIONS WELL INTO THE DIGITAL CAMERA MARKET

George Eastman was born in 1854 in Upstate New York. A high school dropout because of the untimely death of his father and the ensuing financial needs of his family, Eastman began his business career as a 14-year-old office boy in an insurance company, eventually taking charge of policy filing and even policy writing. In a few years, after studying accounting on the side, he became a bank clerk. When Eastman was 24, he was making plans for a trip to the Caribbean when a coworker suggested that he record the trip. Eastman bought considerable photographic equipment and learned how to take pictures. He described the complete outfit of his photographic equipment as a "pack-horse load." Eastman never made the trip to the Caribbean, but he did spend the rest of his life in pursuit of simplifying the complicated photographic process, reducing the size of the load, and making picture taking available to everyone.

After experimenting for several years with various chemical procedures related to the photographic process at night while maintaining his job at the bank during the day, Eastman found a dry plate formula that worked, and he patented a machine for preparing large numbers of dry plates. In April 1880, he leased the third floor of a building in Rochester, New York, and began to manufacture dry plates for sale. Soon Eastman realized that what he was actually doing was attempting to make photography an everyday affair. He described it as trying "to make the camera as convenient as the pencil." In Eastman's experiments, he discovered how to use paper with a layer of plain, soluble gelatin mounted on a roll holder to take pictures rather than using dry plates. In 1885, he began mass advertising his products, and 1888, he introduced the Kodak camera. By 1896, 100,000 Kodak cameras had been produced, and film and photographic paper was being made at a rate of about 400 miles a month. Today, the trademark, "Kodak," coined by Eastman, himself, is known around the world for excellence in photographic products. Kodak has manufacturing operations in North America, South America, and Europe, and Kodak products are available in virtually every country in the world. Kodak's products include digital cameras, healthcare scanners, printing equipment, imaging, radiography, projectors, film, digital imaging products, and many more. By the year 2007, the Eastman Kodak company had 26,900 employees and sales of over $ 10.3 billion.

Discussion

Suppose you are a part of a Kodak team whose task it is to examine quality, customer satisfaction, and market issues. Using techniques presented in this chapter, analyze and discuss the following questions:

1. According to general market information, Kodak is number three in the sales of digital cameras in the United States, with a market share of 16%. However, your team wants to confirm that this figure is constant for various geographic segments of the country. In an effort to study this issue, a random sample of 30 current purchasers of digital cameras is taken in each of the Northeast, the South, the Midwest, and the West. If the 16% market share figure is constant across regions, how many of the 30 purchases of digital cameras would the company expect to be Kodak cameras in each region? If 8 or more of the 30 purchases in the Northeast are Kodak, what might that tell the team? If fewer than 3 of the 30 purchases in the South are Kodak brand, what does that mean? Suppose none of the 30 purchases in the Midwest is Kodak brand digital cameras. Is it still possible that Kodak holds 16% of the market share? Why or why not? If, indeed, Kodak holds a 16% share of the market in the West, is it likely that in a sample of 30 purchases that 20 or more are Kodak? Explain to the team.
2. Digital cameras have been quickly replacing film cameras in recent years. Companies that did not respond to the rapid market change were severely hurt and several went out of business. Kodak responded by embracing the new digital picture-taking platform, while at the same time

continuing its efforts in the production and marketing of film and film cameras thereby creating a somewhat seamless segue into the digital market. Suppose the Kodak team wants to ascertain if people take more or fewer pictures with the digital format than they did with film cameras. Suppose in a previous study using film cameras, it was determined that, on average during daylight hours, families on vacation took 1.5 pictures per hour. Using this figure as a guide, if the Kodak team randomly samples families on vacation in various parts of the United States who are using digital cameras to take pictures, what is the probability that a family takes four or more pictures in an hour? What is the probability that a family takes six or more pictures per hour? What is the probability that a family takes nine or more pictures per hour? What might the answers to these questions indicate about the usage of digital cameras versus film cameras?

3. According to company information, Kodak Easyshare Digital Cameras ranked highest in customer satisfaction in the \$200 to \$399 and \$400 to \$599 price segments in a recent year. Suppose a consumer group conducts a study of 60 recent purchasers of digital cameras, of which 14 own a Kodak Easyshare Digital Camera. In the study, camera owners are asked to rate their satisfaction with their cameras on a scale from 0 to 100. The top 10 satisfaction scores are taken, and 4 of the top 10 are from owners of Kodak Easyshare Digital Cameras. Is this about what is to be expected given the number of owners of this camera in the pool of 60 purchasers? If not, how can you explain the disparity? Suppose seven of the top 10 satisfaction scores were obtained from Kodak Easyshare Digital Camera purchasers. What might this indicate?

Adapted from: Information found at Kodak's Web site: http://www.kodak.com and "Kodak Tops USA Digital Camera Market" at Web site: http://www.letsgodigital.org/en/news/articles/story_6315.html. http://www.macworld.com/article/55236/2007/02/cameras.html; http://www.hoovers.com/eastman-kodak/--ID_10500--/free-co-factsheet.xhtml

USING THE COMPUTER

EXCEL

- Excel can be used to compute exact or cumulative probabilities for particular values of discrete distributions including the binomial, Poisson, and hypergeometric distributions.
- Calculation of probabilities from each of these distributions begins with the **Insert Function** (f_x). To access the **Insert Function**, go to the **Formulas** tab on an Excel worksheet (top center tab). The **Insert Function** is on the far left of the menu bar. In the **Insert Function** dialog box at the top, there is a pulldown menu where it says **Or select a category**. From the pulldown menu associated with this command, select **Statistical**.
- To compute probabilities from a binomial distribution, select **BINOMDIST** from the **Insert Function**'s **Statistical** menu. In the **BINOMDIST** dialog box, there are four lines to which you must respond. On the first line, **Number_s**, enter the value of x, the number of successes. On the second line, **Trials**, enter the number of trials (sample size, n). On the third line, **Probability_s**, enter the value of p. The fourth line, **Cumulative**, requires a logical response of either TRUE or FALSE. Place TRUE in the slot to get the cumulative probabilities for all values from 0 to x. Place FALSE in the slot to get the exact probability of getting x successes in n trials.
- To compute probabilities from a Poisson distribution, select **POISSON** from the **Insert Function**'s **Statistical** menu. In the **POISSON** dialog box, there are three lines to which you must respond. On the first line, **X**, enter the value of x, the number of events. On the second line, **Mean**, enter the expected number, λ. The third line, **Cumulative**, requires a logical response of either TRUE or FALSE. Place TRUE in the slot to get the cumulative probabilities for all values from 0 to x. Place FALSE in the slot to get the exact probability of getting x successes when λ is the expected number.
- To compute probabilities from a hypergeometric distribution, select **HYPGEOMDIST** from the **Insert Function**'s **Statistical** menu. In the **HYPGEOMDIST** dialog box, there are four lines to which you must respond. On the first line, **Sample_s**, enter the value of x, the number of successes in the sample. On the second line, **Number_sample**, enter the size of the sample, n. On the third line, **Population_s**, enter the number of successes in the population. The fourth line, **Number_pop**, enter the size of the population, N.

MINITAB

- Probabilities can be computed using Minitab for the binomial distribution, the Poisson distribution, and the hypergeometric distribution.
- To begin binomial distribution probabilities, select **Calc** on the menu bar. Select **Probability Distributions** from the pulldown menu. From the long second pulldown menu, select **Binomial**. From the dialog box, check how you want the probabilities to be calculated from **Probability**, **Cumulative probability**, or **Inverse cumulative probability**. **Probability** yields the exact probability n, p, and x. **Cumulative probability** produces the cumulative probabilities for values less than or equal to x. **Inverse probability** yields the inverse of the cumulative probabilities. If you want to compute probabilities for several values of x, place them

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

- To begin Poisson distribution probabilities, select **Calc** on the menu bar. Select **Probability Distributions** from the pulldown menu. From the long second pulldown menu, select **Poisson**. From the dialog box, check how you want the probabilities to be calculated from **Probability**, **Cumulative probability**, or **Inverse cumulative probability**. **Probability** yields the exact probability of a particular λ, and x. **Cumulative probability** produces the cumulative probabilities for values less than or equal to x. **Inverse probability** yields the inverse of the cumulative probabilities. If you want to compute probabilities for several values of x, place them in a column, and list the column location in **Input column**. If you want to compute the probability for a particular value of x, check **Input constant**, and enter the value of x.
- To begin hypergeometric distribution probabilities, select **Calc** on the menu bar. Select **Probability Distributions** from the pulldown menu. From the long second pull-down menu, select **Hypergeometric**. From the dialog box, check how you want the probabilities to be calculated from **Probability**, **Cumulative probability**, or **Inverse cumulative probability**. **Probability** yields the exact probability of a particular combination of N, A, n, and x. Note that Minitab uses M for number of successes in the population instead of A. **Cumulative probability** produces the cumulative probabilities for values less than or equal to x. **Inverse probability** yields the inverse of the cumulative probabilities. If you want to compute probabilities for several values of x, place them in a column, and list the column location in **Input column**. If you want to compute the probability for a particular value of x, check **Input constant**, and enter the value of x.