© Comstock Images/Jupiterimages

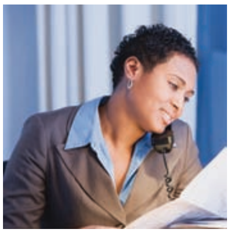
RANDOM VARIABLES AND DISCRETE PROBABILITY DISTRIBUTIONS

Investing to Maximize Returns and Minimize Risk

DATA Xm07-00

An investor has $100,000 to invest in the stock market. She is interested in developing a stock portfolio made up of stocks on the New York Stock Exchange (NYSE), the Toronto Stock Exchange (TSX), and the NASDAQ. The stocks are Coca Cola and Disney (NYSE), Barrick Gold (TSX), and Amazon (NASDAQ). However, she doesn't know how much to invest in each one. She wants to maximize her return, but she would also like to minimize the risk. She has computed the monthly returns for all four stocks during a 60-month period (January 2005 to December 2009). After some consideration, she narrowed her choices down to the following three. What should she do?

1. $25,000 in each stock
2. Coca Cola: $10,000, Disney: $20,000, Barrick Gold: $30,000, Amazon: $40,000
3. Coca Cola: $10,000, Disney: $50,000, Barrick Gold: $30,000, Amazon: $10,000



© Terry Vine/Blend Images/ Jupiterimages

We will provide our answer after we've developed the necessary tools in Section 7.3.

(Case 6.2 continued)

is on first base is .39. If the bases are loaded with one out, then the probability of scoring any runs is .67.

TABLE 1 Probability of Scoring Any Runs

Bases Occupied	0 Outs	1 Out	2 Outs
Bases empty	.26	.16	.07
First base	.39	.26	.13
Second base	.57	.42	.24
Third base	.72	.55	.28
First base and second base	.59	.45	.24
First base and third base	.76	.61	.37
Second base and third base	.83	.74	.37
Bases loaded	.81	.67	.43

(Probabilities are based on results from the American League during the 1989 season. The results for the National League are also shown in the article and are similar.)

Table 1 allows us to determine the best strategy in a variety of circumstances. This case will concentrate on the strategy of the sacrifice bunt. The purpose of the sacrifice bunt is to sacrifice the batter to move base runners to the next base. It can be employed when there are fewer than two outs and men on base. Ignoring the suicide squeeze, any of four outcomes can occur:

1. The bunt is successful. The runner (or runners) advances one base, and the batter is out.
2. The batter is out but fails to advance the runner.
3. The batter bunts into a double play.
4. The batter is safe (hit or error), and the runner advances.

Suppose that you are an American League manager. The game is tied in the middle innings of a game, and there is a runner on first base with no one out. Given the following probabilities of the four outcomes of a bunt for the batter at the plate, should you signal the batter to sacrifice bunt?

$$P(\text{Outcome } 1) = .75$$
$$P(\text{Outcome } 2) = .10$$
$$P(\text{Outcome } 3) = .10$$
$$P(\text{Outcome } 4) = .05$$

Assume for simplicity that after the hit or error in outcome 4, there will be men on first and second base and no one out.

CASE 6.3 Should He Attempt to Steal a Base?

© AP Photo/Al Behrman

Refer to Case 6.2. Another baseball strategy is to attempt to steal second base. Historically the probability of a successful steal of second base is approximately 68%. The probability of being thrown out is 32%. (We'll ignore the relatively rare event wherein the catcher throws the ball into center field allowing the base runner to advance to third base.) Suppose there is a runner on first base. For each of the possible number of outs (0, 1, or 2), determine whether it is advisable to have the runner attempt to steal second base.

CASE 6.4 Maternal Serum Screening Test for Down Syndrome

© AP Photo/Javier Galeano

Pregnant women are screened for a birth defect called Down syndrome. Down syndrome babies are mentally and physically challenged. Some mothers choose to abort the fetus when they are certain that their baby will be born with the syndrome. The most common screening is maternal serum screening, a blood test that looks for markers in the blood to indicate whether the birth defect may occur. The false-positive and false-negative rates vary according to the age of the mother.

This page intentionally left blank

INTRODUCTION

In this chapter, we extend the concepts and techniques of probability introduced in Chapter 6. We present random variables and probability distributions, which are essential in the development of statistical inference.

Here is a brief glimpse into the wonderful world of statistical inference. Suppose that you flip a coin 100 times and count the number of heads. The objective is to determine whether we can infer from the count that the coin is not balanced. It is reasonable to believe that observing a large number of heads (say, 90) or a small number (say, 15) would be a statistical indication of an unbalanced coin. However, where do we draw the line? At 75 or 65 or 55? Without knowing the probability of the frequency of the number of heads from a balanced coin, we cannot draw any conclusions from the sample of 100 coin flips.

The concepts and techniques of probability introduced in this chapter will allow us to calculate the probability we seek. As a first step, we introduce random variables and probability distributions.

7.1 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Consider an experiment where we flip two balanced coins and observe the results. We can represent the events as

Heads on the first coin and heads on the second coin

Heads on the first coin and tails on the second coin

Tails on the first coin and heads on the second coin

Tails on the first coin and tails on the second coin

However, we can list the events in a different way. Instead of defining the events by describing the outcome of each coin, we can count the number of heads (or, if we wish, the number of tails). Thus, the events are now

2 heads

1 heads

1 heads

0 heads

The number of heads is called the **random variable**. We often label the random variable X, and we're interested in the probability of each value of X. Thus, in this illustration, the values of X are 0, 1, and 2.

Here is another example. In many parlor games as well as in the game of craps played in casinos, the player tosses two dice. One way of listing the events is to describe the number on the first die and the number on the second die as follows.

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

However, in almost all games, the player is primarily interested in the total. Accordingly, we can list the totals of the two dice instead of the individual numbers.

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

If we define the random variable X as the total of the two dice, then X can equal 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12.

Random Variable

A **random variable** is a function or rule that assigns a number to each outcome of an experiment.

In some experiments the outcomes are numbers. For example, when we observe the return on an investment or measure the amount of time to assemble a computer, the experiment produces events that are numbers. Simply stated, the value of a random variable is a numerical event.

There are two types of random variables, discrete and continuous. A **discrete random variable** is one that can take on a countable number of values. For example, if we define X as the number of heads observed in an experiment that flips a coin 10 times, then the values of X are 0, 1, 2, . . . , 10. The variable X can assume a total of 11 values. Obviously, we counted the number of values; hence, X is discrete.

A **continuous random variable** is one whose values are uncountable. An excellent example of a continuous random variable is the amount of time to complete a task. For example, let X = time to write a statistics exam in a university where the time limit is 3 hours and students cannot leave before 30 minutes. The smallest value of X is 30 minutes. If we attempt to count the number of values that X can take on, we need to identify the next value. Is it 30.1 minutes? 30.01 minutes? 30.001 minutes? None of these is the second possible value of X because there exist numbers larger than 30 and smaller than 30.001. It becomes clear that we cannot identify the second, or third, or any other values of X (except for the largest value 180 minutes). Thus, we cannot count the number of values, and X is continuous.

A **probability distribution** is a table, formula, or graph that describes the values of a random variable and the probability associated with these values. We will address discrete probability distributions in the rest of this chapter and cover continuous distributions in Chapter 8.

As we noted above, an uppercase letter will represent the *name* of the random variable, usually X. Its lowercase counterpart will represent the value of the random variable. Thus, we represent the probability that the random variable X will equal x as

$$P(X = x)$$

or more simply

$$P(x)$$

Discrete Probability Distributions

The probabilities of the values of a discrete random variable may be derived by means of probability tools such as tree diagrams or by applying one of the definitions of probability. However, two fundamental requirements apply as stated in the box.

Requirements for a Distribution of a Discrete Random Variable

1. $0 \leq P(x) \leq 1 \quad$ for all x
2. $\sum_{\text{all } x} P(x) = 1$

where the random variable can assume values x and $P(x)$ is the probability that the random variable is equal to x.

These requirements are equivalent to the rules of probability provided in Chapter 6. To illustrate, consider the following example.

EXAMPLE 7.1

Probability Distribution of Persons per Household

The *Statistical Abstract of the United States* is published annually. It contains a wide variety of information based on the census as well as other sources. The objective is to provide information about a variety of different aspects of the lives of the country's residents. One of the questions asks households to report the number of persons living in the household. The following table summarizes the data. Develop the probability distribution of the random variable defined as the number of persons per household.

Number of Persons	Number of Households (Millions)
1	31.1
2	38.6
3	18.8
4	16.2
5	7.2
6	2.7
7 or more	1.4
Total	116.0

Source: *Statistical Abstract of the United States, 2009*, Table 61.

SOLUTION

The probability of each value of X, the number of persons per household is computed as the relative frequency. We divide the frequency for each value of X by the total number of households, producing the following probability distribution.

x	$P(x)$
1	31.1/116.0 = .268
2	38.6/116.0 = .333
3	18.8/116.0 = .162
4	16.2/116.0 = .140
5	7.2/116.0 = .062
6	2.7/116.0 = .023
7 or more	1.4/116.0 = .012
Total	1.000

As you can see, the requirements are satisfied. Each probability lies between 0 and 1, and the total is 1.

We interpret the probabilities in the same way we did in Chapter 6. For example, if we select one household at random, the probability that it has three persons is

$$P(3) = .162$$

We can also apply the addition rule for mutually exclusive events. (The values of X are mutually exclusive; a household can have 1, 2, 3, 4, 5, 6, or 7 or more persons.) The probability that a randomly selected household has four or more persons is

$$\begin{aligned} P(X \geq 4) &= P(4) + P(5) + P(6) + P(7 \text{ or more}) \\ &= .140 + .062 + .023 + .012 = .237 \end{aligned}$$

In Example 7.1, we calculated the probabilities using census information about the entire population. The next example illustrates the use of the techniques introduced in Chapter 6 to develop a probability distribution.

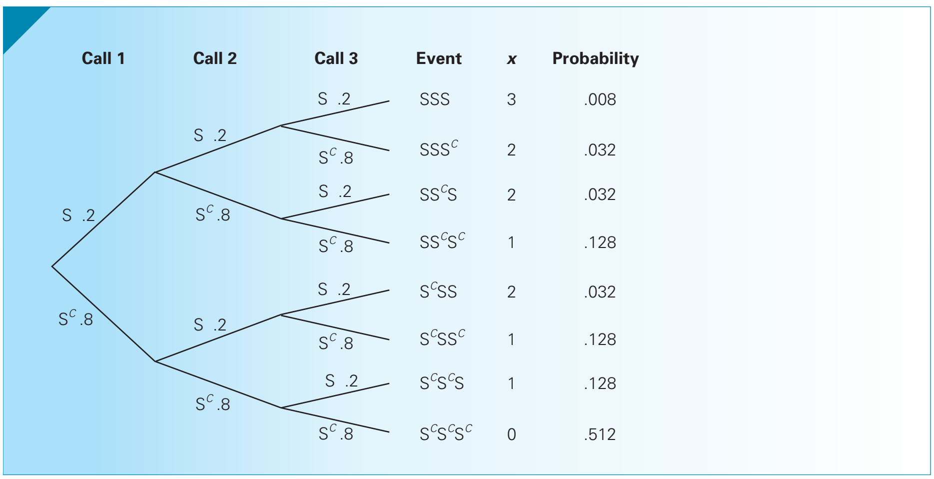
EXAMPLE 7.2 Probability Distribution of the Number of Sales

A mutual fund salesperson has arranged to call on three people tomorrow. Based on past experience, the salesperson knows there is a 20% chance of closing a sale on each call. Determine the probability distribution of the number of sales the salesperson will make.

SOLUTION

We can use the probability rules and trees introduced in Section 6.3. Figure 7.1 displays the probability tree for this example. Let X = the number of sales.

FIGURE 7.1



The tree exhibits each of the eight possible outcomes and their probabilities. We see that there is one outcome that represents no sales, and its probability is $P(0) = .512$. There are three outcomes representing one sale, each with probability .128, so we add these probabilities. Thus,

$$P(1) = .128 + .128 + .128 = 3(.128) = .384$$

The probability of two sales is computed similarly:

$$P(X) = 3(.032) = .096$$

There is one outcome where there are three sales:

$$P(3) = .008$$

The probability distribution of X is listed in Table 7.1.

TABLE **7.1** Probability Distribution of the Number of Sales in Example 7.2

x	$P(x)$
0	.512
1	.384
2	.096
3	.008

Probability Distributions and Populations

The importance of probability distributions derives from their use as representatives of populations. In Example 7.1, the distribution provided us with information about the population of numbers of persons per household. In Example 7.2, the population was the number of sales made in three calls by the salesperson. And as we noted before, statistical inference deals with inference about populations.

Describing the Population/Probability Distribution

In Chapter 4, we showed how to calculate the mean, variance, and standard deviation of a population. The formulas we provided were based on knowing the value of the random variable for each member of the population. For example, if we want to know the mean and variance of annual income of all North American blue-collar workers, we would record each of their incomes and use the formulas introduced in Chapter 4:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

where X_1 is the income of the first blue-collar worker, X_2 is the second worker's income, and so on. It is likely that N equals several million. As you can appreciate, these formulas are seldom used in practical applications because populations are so large. It is unlikely that we would be able to record all the incomes in the population of North American blue-collar workers. However, probability distributions often represent populations. Rather than record each of the many observations in a population, we list the values and their associated probabilities as we did in deriving the probability distribution of the number of persons per household in Example 7.1 and the number of successes in three calls by the mutual fund salesperson. These can be used to compute the mean and variance of the population.

The population mean is the weighted average of all of its values. The weights are the probabilities. This parameter is also called the **expected value** of X and is represented by $E(X)$.

Population Mean

$$E(X) = \mu = \sum_{\text{all } x} xP(x)$$

The population variance is calculated similarly. It is the weighted average of the squared deviations from the mean.

Population Variance

$$V(X) = \sigma^2 = \sum_{\text{all } x} (x - \mu)^2 P(x)$$

There is a shortcut calculation that simplifies the calculations for the population variance. This formula is not an approximation; it will yield the same value as the formula above.

Shortcut Calculation for Population Variance

$$V(X) = \sigma^2 = \sum_{\text{all } x} x^2 P(x) - \mu^2$$

The standard deviation is defined as in Chapter 4.

Population Standard Deviation

$$\sigma = \sqrt{\sigma^2}$$

EXAMPLE 7.3

Describing the Population of the Number of Persons per Household

Find the mean, variance, and standard deviation for the population of the number of persons per household Example 7.1.

SOLUTION

For this example, we will assume that the last category is exactly seven persons. The mean of X is

$$\begin{aligned} E(X) = \mu = \sum_{\text{all } x} xP(x) &= 1P(1) + 2P(2) + 3P(3) + 4P(4) + 5P(5) + 6P(6) + 7P(7) \\ &= 1(.268) + 2(.333) + 3(.162) + 4(.140) + 5(.062) + 6(.023) + 7(.012) \\ &= 2.512 \end{aligned}$$

Notice that the random variable can assume integer values only, yet the mean is 2.513.

The variance of X is

$$\begin{aligned} V(X) = \sigma^2 &= \sum_{\text{all } x} (x - \mu)^2 P(x) \\ &= (1 - 2.512)^2(.268) + (2 - 2.512)^2(.333) + (3 - 2.512)^2(.162) \\ &\quad + (4 - 2.512)^2(.140) + (5 - 2.512)^2(.062) + (6 - 2.512)^2(.023) \\ &\quad + (7 - 2.512)^2(.012) \\ &= 1.954 \end{aligned}$$

To demonstrate the shortcut method, we'll use it to recompute the variance:

$$\begin{aligned} \sum_{\text{all } x} x^2 P(x) &= 1^2(.268) + 2^2(.333) + 3^2(.162) + 4^2(.140) + 5^2(.062) \\ &\quad + 6^2(.023) + 7^2(.012) = 8.264 \end{aligned}$$

and

$$\mu = 2.512$$

Thus,

$$\sigma^2 = \sum_{\text{all } x} x^2 P(x) - \mu^2 = 8.264 - (2.512)^2 = 1.954$$

The standard deviation is

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.954} = 1.398$$

These parameters tell us that the mean and standard deviation of the number of persons per household are 2.512 and 1.398, respectively.

Laws of Expected Value and Variance

As you will discover, we often create new variables that are functions of other random variables. The formulas given in the next two boxes allow us to quickly determine the expected value and variance of these new variables. In the notation used here, X is the random variable and c is a constant.

Laws of Expected Value

1. $E(c) = c$
2. $E(X + c) = E(X) + c$
3. $E(cX) = cE(X)$

Laws of Variance

1. $V(c) = 0$
2. $V(X + c) = V(X)$
3. $V(cX) = c^2V(X)$

EXAMPLE 7.4

Describing the Population of Monthly Profits

The monthly sales at a computer store have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000. Find the mean and standard deviation of monthly profits.

SOLUTION

We can describe the relationship between profits and sales by the following equation:

$$\text{Profit} = .30(\text{Sales}) - 6{,}000$$

The expected or mean profit is

$$E(\text{Profit}) = E[.30(\text{Sales}) - 6{,}000]$$

Applying the second law of expected value, we produce

$$E(\text{Profit}) = E[.30(\text{Sales})] - 6{,}000$$

Applying law 3 yields

$$E(\text{Profit}) = .30E(\text{Sales}) - 6{,}000 = .30(25{,}000) - 6{,}000 = 1{,}500$$

Thus, the mean monthly profit is \$1,500.

The variance is

$$V(\text{Profit}) = V[.30(\text{Sales}) - 6{,}000]$$

The second law of variance states that

$$V(\text{Profit}) = V[.30(\text{Sales})]$$

and law 3 yields

$$V(\text{Profit}) = (.30)^2V(\text{Sales}) = .09(4{,}000)^2 = 1{,}440{,}000$$

Thus, the standard deviation of monthly profits is

$$\sigma_{\text{Profit}} = \sqrt{1{,}440{,}000} = \$1{,}200$$

EXERCISES

7.1 The number of accidents that occur on a busy stretch of highway is a random variable.

a. What are the possible values of this random variable?
b. Are the values countable? Explain.
c. Is there a finite number of values? Explain.
d. Is the random variable discrete or continuous? Explain.

7.2 The distance a car travels on a tank of gasoline is a random variable.

a. What are the possible values of this random variable?
b. Are the values countable? Explain.
c. Is there a finite number of values? Explain.
d. Is the random variable discrete or continuous? Explain.

7.3 The amount of money students earn on their summer jobs is a random variable.

a. What are the possible values of this random variable?
b. Are the values countable? Explain.
c. Is there a finite number of values? Explain.
d. Is the random variable discrete or continuous? Explain.

7.4 The mark on a statistics exam that consists of 100 multiple-choice questions is a random variable.

a. What are the possible values of this random variable?
b. Are the values countable? Explain.
c. Is there a finite number of values? Explain.
d. Is the random variable discrete or continuous? Explain.

7.5 Determine whether each of the following is a valid probability distribution.

a.

x	0	1	2	3
$P(x)$	.1	.3	.4	.1

b.

x	5	−6	10	0
$P(x)$	.01	.01	.01	.97

c.

x	14	12	−7	13
$P(x)$	.25	.46	.04	.24

7.6 Let X be the random variable designating the number of spots that turn up when a balanced die is rolled. What is the probability distribution of X?

7.7 In a recent census the number of color televisions per household was recorded

Number of color televisions	0	1	2	3	4	5
Number of households (thousands)	1,218	32,379	37,961	19,387	7,714	2,842

a. Develop the probability distribution of X, the number of color televisions per household.
b. Determine the following probabilities.

$P(X \le 2)$
$P(X > 2)$
$P(X \ge 4)$

7.8 Using historical records, the personnel manager of a plant has determined the probability distribution of X, the number of employees absent per day. It is

x	0	1	2	3	4	5	6	7
$P(x)$	.005	.025	.310	.340	.220	.080	.019	.001

a. Find the following probabilities.

$P(2 \le X \le 5)$
$P(X > 5)$
$P(X < 4)$

b. Calculate the mean of the population.
c. Calculate the standard deviation of the population.

7.9 Second-year business students at many universities are required to take 10 one-semester courses. The number of courses that result in a grade of A is a discrete random variable. Suppose that each value of this random variable has the same probability. Determine the probability distribution.

7.10 The random variable X has the following probability distribution.

x	−3	2	6	8
$P(x)$	.2	.3	.4	.1

Find the following probabilities.

a. $P(X > 0)$
b. $P(X \ge 1)$
c. $P(X \ge 2)$
d. $P(2 \le X \le 5)$

7.11 An Internet pharmacy advertises that it will deliver the over-the-counter products that customers purchase in 3 to 6 days. The manager of the company wanted to be more precise in its advertising. Accordingly, she recorded the number of days it took to deliver to customers. From the data, the following probability distribution was developed.

Number of days	0	1	2	3	4	5	6	7	8
Probability	0	0	.01	.04	.28	.42	.21	.02	.02

a. What is the probability that a delivery will be made within the advertised 3- to 6-day period?

b. What is the probability that a delivery will be late?
c. What is the probability that a delivery will be early?

7.12 A gambler believes that a strategy called "doubling up" is an effective way to gamble. The method requires the gambler to double the stake after each loss. Thus, if the initial bet is $1, after losing he will double the bet until he wins. After a win, he resorts back to a $1 bet. The result is that he will net $1 for every win. The problem however, is that he will eventually run out of money or bump up against the table limit. Suppose that for a certain game the probability of winning is .5 and that losing six in a row will result in bankrupting the gambler. Find the probability of losing six times in a row.

7.13 The probability that a university graduate will be offered no jobs within a month of graduation is estimated to be 5%. The probability of receiving one, two, and three job offers has similarly been estimated to be 43%, 31%, and 21%, respectively. Determine the following probabilities.
a. A graduate is offered fewer than two jobs.
b. A graduate is offered more than one job.

7.14 Use a probability tree to compute the probability of the following events when flipping two fair coins.
a. Heads on the first coin and heads on the second coin
b. Heads on the first coin and tails on the second coin
c. Tails on the first coin and heads on the second coin
d. Tails on the first coin and tails on the second coin

7.15 Refer to Exercise 7.14. Find the following probabilities.
a. No heads
b. One head
c. Two heads
d. At least one head

7.16 Draw a probability tree to describe the flipping of three fair coins.

7.17 Refer to Exercise 7.16. Find the following probabilities.
a. Two heads
b. One head
c. At least one head
d. At least two heads

7.18 The random variable X has the following distribution.

x	-2	5	7	8
$P(x)$	.59	.15	.25	.01

a. Find the mean and variance for the probability distribution below.
b. Determine the probability distribution of Y where $Y = 5X$.
c. Use the probability distribution in part (b) to compute the mean and variance of Y.
d. Use the laws of expected value and variance to find the expected value and variance of Y from the parameters of X.

7.19 We are given the following probability distribution.

x	0	1	2	3
$P(x)$	.4	.3	.2	.1

a. Calculate the mean, variance, and standard deviation.
b. Suppose that $Y = 3X + 2$. For each value of X, determine the value of Y. What is the probability distribution of Y?
c. Calculate the mean, variance, and standard deviation from the probability distribution of Y.
d. Use the laws of expected value and variance to calculate the mean, variance, and standard deviation of Y from the mean, variance, and standard deviation of X. Compare your answers in parts (c) and (d). Are they the same (except for rounding)?

7.20 The number of pizzas delivered to university students each month is a random variable with the following probability distribution.

x	0	1	2	3
$P(X)$	.1	.3	.4	.2

a. Find the probability that a student has received delivery of two or more pizzas this month.
b. Determine the mean and variance of the number of pizzas delivered to students each month.

7.21 Refer to Exercise 7.20. If the pizzeria makes a profit of $3 per pizza, determine the mean and variance of the profits per student.

7.22 After watching a number of children playing games at a video arcade, a statistics practitioner estimated the following probability distribution of X, the number of games per visit.

x	1	2	3	4	5	6	7
$P(x)$	.05	.15	.15	.25	.20	.10	.10

a. What is the probability that a child will play more than four games?
b. What is the probability that a child will play at least two games?

7.23 Refer to Exercise 7.22. Determine the mean and variance of the number of games played.

7.24 Refer to Exercise 7.23. Suppose that each game costs the player 25 cents. Use the laws of expected value and variance to determine the expected value and variance of the amount of money the arcade takes in.

7.25 Refer to Exercise 7.22.
a. Determine the probability distribution of the amount of money the arcade takes in per child.
b. Use the probability distribution to calculate the mean and variance of the amount of money the arcade takes in.
c. Compare the answers in part (b) with those of Exercise 7.24. Are they identical (except for rounding errors)?

7.26 A survey of Amazon.com shoppers reveals the following probability distribution of the number of books purchased per hit.

x	0	1	2	3	4	5	6	7
$P(x)$	.35	.25	.20	.08	.06	.03	.02	.01

a. What is the probability that an Amazon.com visitor will buy four books?
b. What is the probability that an Amazon.com visitor will buy eight books?
c. What is the probability that an Amazon.com visitor will not buy any books?
d. What is the probability that an Amazon.com visitor will buy at least one book?

7.27 A university librarian produced the following probability distribution of the number of times a student walks into the library over the period of a semester.

x	0	5	10	15	20	25	30	40	50	75	100
$P(x)$	.22	.29	.12	.09	.08	.05	.04	.04	.03	.03	.01

Find the following probabilities.
a. $P(X \geq 20)$
b. $P(X = 60)$
c. $P(X > 50)$
d. $P(X > 100)$

7.28 After analyzing the frequency with which cross-country skiers participate in their sport, a sportswriter created the following probability distribution for X = number of times per year cross-country skiers ski.

x	0	1	2	3	4	5	6	7	8
$P(x)$	.04	.09	.19	.21	.16	.12	.08	.06	.05

Find the following.
a. $P(3)$
b. $P(X \geq 5)$
c. $P(5 \leq X \leq 7)$

7.29 The natural remedy echinacea is reputed to boost the immune system, which will reduce the number of flu and colds. A 6-month study was undertaken to determine whether the remedy works. From this study, the following probability distribution of the number of respiratory infections per year (X) for echinacea users was produced.

x	0	1	2	3	4
$P(x)$	.45	.31	.17	.06	.01

Find the following probabilities.
a. An echinacea user has more than one infection per year.
b. An echinacea user has no infections per year.
c. An echinacea user has between one and three (inclusive) infections per year.

7.30 A shopping mall estimates the probability distribution of the number of stores mall customers actually enter, as shown in the table.

x	0	1	2	3	4	5	6
$P(x)$	.04	.19	.22	.28	.12	.09	.06

Find the mean and standard deviation of the number of stores entered.

7.31 Refer to Exercise 7.30. Suppose that, on average, customers spend 10 minutes in each store they enter. Find the mean and standard deviation of the total amount of time customers spend in stores.

7.32 When parking a car in a downtown parking lot, drivers pay according to the number of hours or parts thereof. The probability distribution of the number of hours cars are parked has been estimated as follows.

x	1	2	3	4	5	6	7	8
$P(x)$	.24	.18	.13	.10	.07	.04	.04	.20

Find the mean and standard deviation of the number of hours cars are parked in the lot.

7.33 Refer to Exercise 7.32. The cost of parking is $2.50 per hour. Calculate the mean and standard deviation of the amount of revenue each car generates.

7.34 You have been given the choice of receiving $500 in cash or receiving a gold coin that has a face value of $100. However, the actual value of the gold coin depends on its gold content. You are told that the coin has a 40% probability of being worth $400, a 30% probability of being worth $900, and a 30% probability of being worth its face value. Basing your decision on expected value, should you choose the coin?

7.35 The manager of a bookstore recorded the number of customers who arrive at a checkout counter every 5 minutes from which the following distribution was calculated. Calculate the mean and standard deviation of the random variable.

x	0	1	2	3	4
$P(x)$	.10	.20	.25	.25	.20

7.36 The owner of a small firm has just purchased a personal computer, which she expects will serve her for the next 2 years. The owner has been told that she "must" buy a surge suppressor to provide protection for her new hardware against possible surges or variations in the electrical current, which have the capacity to damage the computer. The amount of damage to the computer depends on the strength of the surge. It has been estimated that there is a 1% chance of incurring $400 damage, a 2% chance of incurring $200 damage, and 10% chance of $100 damage. An inexpensive suppressor, which would provide protection for only one surge can be purchased. How much should the owner be willing to pay if she makes decisions on the basis of expected value?

7.37 It cost one dollar to buy a lottery ticket, which has five prizes. The prizes and the probability that a player wins the prize are listed here. Calculate the expected value of the payoff.

Prize ($)	1 million	200,000	50,000
Probability	1/10 million	1/1 million	1/500,000

Prize ($)	10,000	1,000
Probability	1/50,000	1/10,000

7.38 After an analysis of incoming faxes the manager of an accounting firm determined the probability distribution of the number of pages per facsimile as follows:

x	1	2	3	4	5	6	7
$P(x)$	.05	.12	.20	.30	.15	.10	.08

Compute the mean and variance of the number of pages per fax.

7.39 Refer to Exercise 7.38. Further analysis by the manager revealed that the cost of processing each page of a fax is $.25. Determine the mean and variance of the cost per fax.

7.40 To examine the effectiveness of its four annual advertising promotions, a mail-order company has sent a questionnaire to each of its customers, asking how many of the previous year's promotions prompted orders that would not otherwise have been made. The table lists the probabilities that were derived from the questionnaire, where X is the random variable representing the number of promotions that prompted orders. If we assume that overall customer behavior next year will be the same as last year, what is the expected number of promotions that each customer will take advantage of next year by ordering goods that otherwise would not be purchased?

x	0	1	2	3	4
$P(x)$	.10	.25	.40	.20	.05

7.41 Refer to Exercise 7.40. A previous analysis of historical records found that the mean value of orders for promotional goods is $20, with the company earning a gross profit of 20% on each order. Calculate the expected value of the profit contribution next year.

7.42 Refer to Exercises 7.40 and 7.41. The fixed cost of conducting the four promotions is estimated to be $15,000, with a variable cost of $3.00 per customer for mailing and handling costs. How large a customer base does the company need to cover the cost of promotions?

7.2 BIVARIATE DISTRIBUTIONS

Thus far, we have dealt with the distribution of a *single* variable. However, there are circumstances where we need to know about the relationship between two variables. Recall that we have addressed this problem statistically in Chapter 3 by drawing the scatter diagram and in Chapter 4 by calculating the covariance and the coefficient of correlation. In this section, we present the **bivariate distribution**, which provides probabilities of combinations of two variables. Incidentally, when we need to distinguish between the bivariate distributions and the distributions of one variable, we'll refer to the latter as *univariate* distributions.

The joint probability that two variables will assume the values x and y is denoted $P(x, y)$. A bivariate (or joint) probability distribution of X and Y is a table or formula that lists the joint probabilities for all pairs of values of x and y. As was the case with univariate distributions, the joint probability must satisfy two requirements.

Requirements for a Discrete Bivariate Distribution

1. $0 \le P(x, y) \le 1$ for all pairs of values (x, y)
2. $\sum_{\text{all } x} \sum_{\text{all } y} P(x, y) = 1$

EXAMPLE 7.5

Bivariate Distribution of the Number of House Sales

Xavier and Yvette are real estate agents. Let X denote the number of houses that Xavier will sell in a month and let Y denote the number of houses Yvette will sell in a month. An analysis of their past monthly performances has the following joint probabilities.

Bivariate Probability Distribution

		X		
		0	1	2
	0	.12	.42	.06
Y	1	.21	.06	.03
	2	.07	.02	.01

We interpret these joint probabilities in the same way we did in Chapter 6. For example, the probability that Xavier sells 0 houses and Yvette sells 1 house in the month is $P(0, 1) = .21$.

Marginal Probabilities

As we did in Chapter 6, we can calculate the marginal probabilities by summing across rows or down columns.

Marginal Probability Distribution of X in Example 7.5

$$P(X = 0) = P(0, 0) + P(0, 1) + P(0, 2) = .12 + .21 + .07 = .4$$
$$P(X = 1) = P(1, 0) + P(1, 1) + P(1, 2) = .42 + .06 + .02 = .5$$
$$P(X = 2) = P(2, 0) + P(2, 1) + P(2, 2) = .06 + .03 + .01 = .1$$

The marginal probability distribution of X is

x	$P(x)$
0	.4
1	.5
2	.1

Marginal Probability Distribution of Y in Example 7.5

$$P(Y = 0) = P(0, 0) + P(1, 0) + P(2, 0) = .12 + .42 + .06 = .6$$
$$P(Y = 1) = P(0, 1) + P(1, 1) + P(2, 1) = .21 + .06 + .03 = .3$$
$$P(Y = 2) = P(0, 2) + P(1, 2) + P(2, 2) = .07 + .02 + .01 = .1$$

The marginal probability distribution of Y is

y	$P(y)$
0	.6
1	.3
2	.1

Notice that both marginal probability distributions meet the requirements; the probabilities are between 0 and 1, and they add to 1.

Describing the Bivariate Distribution

As we did with the univariate distribution, we often describe the bivariate distribution by computing the mean, variance, and standard deviation of each variable. We do so by utilizing the marginal probabilities.

Expected Value, Variance, and Standard Deviation of X in Example 7.5

$$E(X) = \mu_X = \sum xP(x) = 0(.4) + 1(.5) + 2(.1) = .7$$

$$V(X) = \sigma_X^2 = \sum (x - \mu_X)^2 P(x) = (0 - .7)^2(.4) + (1 - .7)^2(.5) + (2 - .7)^2(.1) = .41$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{.41} = .64$$

Expected Value, Variance, and Standard Deviation of Y in Example 7.5

$$E(Y) = \mu_Y = \sum yP(y) = 0(.6) + 1(.3) + 2(.1) = .5$$

$$V(Y) = \sigma_Y^2 = \sum (y - \mu_Y)^2 P(y) = (0 - .5)^2(.6) + (1 - .5)^2(.3) + (2 - .5)^2(.1) = .45$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{.45} = .67$$

There are two more parameters we can and need to compute. Both deal with the relationship between the two variables. They are the covariance and the coefficient of correlation. Recall that both were introduced in Chapter 4, where the formulas were based on the assumption that we knew each of the N observations of the population. In this chapter, we compute parameters like the covariance and the coefficient of correlation from the bivariate distribution.

Covariance

The covariance of two discrete variables is defined as

$$\text{COV}(X,Y) = \sigma_{xy} = \sum_{\text{all } x} \sum_{\text{all } y} (x - \mu_X)(y - \mu_Y)P(x, y)$$

Notice that we multiply the deviations from the mean for both X and Y and then multiply by the joint probability.

The calculations are simplified by the following shortcut method.

Shortcut Calculation for Covariance

$$COV(X,Y) = \sigma_{xy} = \sum_{\text{all } x} \sum_{\text{all } y} xyP(x, y) - \mu_X\mu_Y$$

The coefficient of correlation is calculated in the same way as in Chapter 4.

Coefficient of Correlation

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

EXAMPLE 7.6

Describing the Bivariate Distribution

Compute the covariance and the coefficient of correlation between the numbers of houses sold by the two agents in Example 7.5.

SOLUTION

We start by computing the covariance.

$$\begin{aligned}\sigma_{xy} &= \sum_{\text{all } x} \sum_{\text{all } y} (x - \mu_X)(y - \mu_Y)P(x, y)\\ &= (0 - .7)(0 - .5)(.12) + (1 - .7)(0 - .5)(.42) + (2 - .7)(0 - .5)(.06)\\ &\quad + (0 - .7)(1 - .5)(.21) + (1 - .7)(1 - .5)(.06) + (2 - .7)(1 - .5)(03)\\ &\quad + (0 - .7)(2 - .5)(.07) + (1 - .7)(2 - .5)(.02) + (2 - .7)(2 - .5)(.01)\\ &= -.15\end{aligned}$$

As we did with the shortcut method for the variance, we'll recalculate the covariance using its shortcut method.

$$\begin{aligned}\sum_{\text{all } x} \sum_{\text{all } y} xyP(x, y) &= (0)(0)(.12) + (1)(0)(.42) + (2)(0)(.06)\\ &\quad + (0)(1)(.21) + (1)(1)(.06) + (2)(1)(.03)\\ &\quad + (0)(2)(.07) + (1)(2)(.02) + (2)(2)(.01)\\ &= .2\end{aligned}$$

Using the expected values computed above we find

$$\sigma_{xy} = \sum_{\text{all } x} \sum_{\text{all } y} xyP(x, y) - \mu_X\mu_Y = .2 - (.7)(.5) = -.15$$

We also computed the standard deviations above. Thus, the coefficient of correlation is

$$\rho = \frac{\sigma_{xy}}{\sigma_X \sigma_Y} = \frac{-.15}{(.64)(.67)} = -.35$$

There is a weak negative relationship between the two variables: the number of houses Xavier will sell in a month (X) and the number of houses Yvette will sell in a month (Y).

Sum of Two Variables

The bivariate distribution allows us to develop the probability distribution of any combination of the two variables. Of particular interest to us is the sum of two variables. The analysis of this type of distribution leads to an important statistical application in finance, which we present in the next section.

To see how to develop the probability distribution of the sum of two variables from their bivariate distribution, return to Example 7.5. The sum of the two variables X and Y is the total number of houses sold per month. The possible values of $X + Y$ are 0, 1, 2, 3, and 4. The probability that $X + Y = 2$, for example, is obtained by summing the joint probabilities of all pairs of values of X and Y that sum to 2:

$$P(X + Y = 2) = P(0,2) + P(1,1) + P(2,0) = .07 + .06 + .06 = .19$$

We calculate the probabilities of the other values of $X + Y$ similarly, producing the following table.

Probability Distribution of $X + Y$ in Example 7.5

$x + y$	0	1	2	3	4
$P(x + y)$	.12	.63	.19	.05	.01

We can compute the expected value, variance, and standard deviation of $X + Y$ in the usual way.

$$E(X + Y) = 0(.12) + 1(.63) + 2(.19) + 3(.05) + 4(.01) = 1.2$$

$$\begin{aligned} V(X + Y) = \sigma^2_{X+Y} &= (0 - 1.2)^2(.12) + (1 - 1.2)^2(.63) + (2 - 1.2)^2(.19) \\ &\quad + (3 - 1.2)^2(.05) + (4 - 1.2)^2(.01) \\ &= .56 \end{aligned}$$

$$\sigma_{X+Y} = \sqrt{.56} = .75$$

We can derive a number of laws that enable us to compute the expected value and variance of the sum of two variables.

Laws of Expected Value and Variance of the Sum of Two Variables

1. $E(X + Y) = E(X) + E(Y)$
2. $V(X + Y) = V(X) + V(Y) + 2\text{COV}(X, Y)$

If X and Y are independent, $\text{COV}(X, Y) = 0$ and thus $V(X + Y) = V(X) + V(Y)$

EXAMPLE 7.7 Describing the Population of the Total Number of House Sales

Use the rules of expected value and variance of the sum of two variables to calculate the mean and variance of the total number of houses sold per month in Example 7.5.

SOLUTION

Using law 1 we compute the expected value of $X + Y$:

$$E(X + Y) = E(X) + E(Y) = .7 + .5 = 1.2$$

which is the same value we produced directly from the probability distribution of $X + Y$.
We apply law 3 to determine the variance:

$$V(X + Y) = V(X) + V(Y) + 2COV(X,Y) = .41 + .45 + 2(-.15) = .56$$

This is the same value we obtained from the probability distribution of $X + Y$.

We will encounter several applications where we need the laws of expected value and variance for the sum of two variables. Additionally, we will demonstrate an important application in operations management where we need the formulas for the expected value and variance of the sum of more than two variables. See Exercises 7.57–7.60.

EXERCISES

7.43 The following table lists the bivariate distribution of X and Y.

	x	
y	**1**	**2**
1	.5	.1
2	.1	.3

a. Find the marginal probability distribution of X.
b. Find the marginal probability distribution of Y.
c. Compute the mean and variance of X.
d. Compute the mean and variance of Y.

7.44 Refer to Exercise 7.43. Compute the covariance and the coefficient of correlation.

7.45 Refer to Exercise 7.43. Use the laws of expected value and variance of the sum of two variables to compute the mean and variance of $X + Y$.

7.46 Refer to Exercise 7.43.
a. Determine the distribution of $X + Y$.
b. Determine the mean and variance of $X + Y$.
c. Does your answer to part (b) equal the answer to Exercise 7.45?

7.47 The bivariate distribution of X and Y is described here.

	x	
y	**1**	**2**
1	.28	.42
2	.12	.18

a. Find the marginal probability distribution of X.
b. Find the marginal probability distribution of Y.
c. Compute the mean and variance of X.
d. Compute the mean and variance of Y.

7.48 Refer to Exercise 7.47. Compute the covariance and the coefficient of correlation.

7.49 Refer to Exercise 7.47. Use the laws of expected value and variance of the sum of two variables to compute the mean and variance of $X + Y$.

7.50 Refer to Exercise 7.47.
a. Determine the distribution of $X + Y$.
b. Determine the mean and variance of $X + Y$.
c. Does your answer to part (b) equal the answer to Exercise 7.49?

7.51 The joint probability distribution of X and Y is shown in the following table.

	x		
y	**1**	**2**	**3**
1	.42	.12	.06
2	.28	.08	.04

a. Determine the marginal distributions of X and Y.
b. Compute the covariance and coefficient of correlation between X and Y.
c. Develop the probability distribution of $X + Y$.

7.52 The following distributions of X and of Y have been developed. If X and Y are independent, determine the joint probability distribution of X and Y.

x	0	1	2
***p*(*x*)**	.6	.3	.1

y	1	2
***p*(*y*)**	.7	.3

7.53 The distributions of X and of Y are described here. If X and Y are independent, determine the joint probability distribution of X and Y.

x	0	1
***P*(*x*)**	.2	.8

y	1	2	3
***P*(*y*)**	.2	.4	.4

7.54 After analyzing several months of sales data, the owner of an appliance store produced the following joint probability distribution of the number of refrigerators and stoves sold daily.

	Refrigerators		
Stoves	**0**	**1**	**2**
0	.08	.14	.12
1	.09	.17	.13
2	.05	.18	.04

a. Find the marginal probability distribution of the number of refrigerators sold daily.
b. Find the marginal probability distribution of the number of stoves sold daily.
c. Compute the mean and variance of the number of refrigerators sold daily.
d. Compute the mean and variance of the number of stoves sold daily.
e. Compute the covariance and the coefficient of correlation.

7.55 Canadians who visit the United States often buy liquor and cigarettes, which are much cheaper in the United States. However, there are limitations. Canadians visiting in the United States for more than 2 days are allowed to bring into Canada one bottle of liquor and one carton of cigarettes. A Canada customs agent has produced the following joint probability distribution of the number of bottles of liquor and the number of cartons of cigarettes imported by Canadians who have visited the United States for 2 or more days.

	Bottles of Liquor	
Cartons of Cigarettes	**0**	**1**
0	.63	.18
1	.09	.10

a. Find the marginal probability distribution of the number of bottles imported.
b. Find the marginal probability distribution of the number of cigarette cartons imported.
c. Compute the mean and variance of the number of bottles imported.
d. Compute the mean and variance of the number of cigarette cartons imported.
e. Compute the covariance and the coefficient of correlation.

7.56 Refer to Exercise 7.54. Find the following conditional probabilities.
a. P(1 refrigerator | 0 stoves)
b. P(0 stoves | 1 refrigerator)
c. P(2 refrigerators | 2 stoves)

APPLICATIONS in OPERATIONS MANAGEMENT

© Susan Van Etten

PERT/CPM

The Project Evaluation and Review Technique (**PERT**) and the Critical Path Method (**CPM**) are related management-science techniques that help operations managers control the activities and the amount of time it takes to complete a project. Both techniques are based on the order in which the activities must be performed. For example, in building a house the excavation of the foundation must precede the pouring of the foundation, which in turn precedes the framing. A **path**

(Continued)

is defined as a sequence of related activities that leads from the starting point to the completion of a project. In most projects, there are several paths with differing amounts of time needed for their completion. The longest path is called the **critical path** because any delay in the activities along this path will result in a delay in the completion of the project. In some versions of PERT/CPM, the activity completion times are fixed and the chief task of the operations manager is to determine the critical path. In other versions, each activity's completion time is considered to be a random variable, where the mean and variance can be estimated. By extending the laws of expected value and variance for the sum of two variables to more than two variables, we produce the following, where $X_1, X_2, \ldots, X_k$ are the times for the completion of activities $1, 2, \ldots, k$, respectively. These times are independent random variables.

Laws of Expected Value and Variance for the Sum of More than Two Independent Variables

1. $E(X_1 + X_2 + \cdots + X_k) = E(X_1) + E(X_2) + \cdots + E(X_k)$
2. $V(X_1 + X_2 + \cdots + X_k) = V(X_1) + V(X_2) + \cdots + V(X_k)$

Using these laws, we can then produce the expected value and variance for the complete project. Exercises 7.57–7.60 address this problem.

7.57 There are four activities along the critical path for a project. The expected values and variances of the completion times of the activities are listed here. Determine the expected value and variance of the completion time of the project.

Activity	Expected Completion Time (Days)	Variance
1	18	8
2	12	5
3	27	6
4	8	2

7.58 The operations manager of a large plant wishes to overhaul a machine. After conducting a PERT/CPM analysis he has developed the following critical path.

1. Disassemble machine
2. Determine parts that need replacing
3. Find needed parts in inventory
4. Reassemble machine
5. Test machine

He has estimated the mean (in minutes) and variances of the completion times as follows.

Activity	Mean	Variance
1	35	8
2	20	5
3	20	4
4	50	12
5	20	2

Determine the mean and variance of the completion time of the project.

7.59 In preparing to launch a new product, a marketing manager has determined the critical path for her department. The activities and the mean and variance of the completion time for each activity along the critical path are shown in the accompanying table. Determine the mean and variance of the completion time of the project.

Activity	Expected Completion Time (Days)	Variance
Develop survey questionnaire	8	2
Pretest the questionnaire	14	5
Revise the questionnaire	5	1
Hire survey company	3	1
Conduct survey	30	8
Analyze data	30	10
Prepare report	10	3

7.60 A professor of business statistics is about to begin work on a new research project. Because his time is quite limited, he has developed a PERT/CPM critical path, which consists of the following activities:

1. Conduct a search for relevant research articles.
2. Write a proposal for a research grant.
3. Perform the analysis.
4. Write the article and send to journal.
5. Wait for reviews.
6. Revise on the basis of the reviews and resubmit.

The mean (in days) and variance of the completion times are as follows

Activity	Mean	Variance
1	10	9
2	3	0
3	30	100
4	5	1
5	100	400
6	20	64

Compute the mean and variance of the completion time of the entire project.

7.3 (Optional) Applications in Finance: Portfolio Diversification and Asset Allocation

In this section we introduce an important application in finance that is based on the previous section.

In Chapter 3 (page 51), we described how the variance or standard deviation can be used to measure the risk associated with an investment. Most investors tend to be risk averse, which means that they prefer to have lower risk associated with their investments. One of the ways in which financial analysts lower the risk that is associated with the stock market is through **diversification**. This strategy was first mathematically developed by Harry Markowitz in 1952. His model paved the way for the development of modern portfolio theory (MPT), which is the concept underlying mutual funds (see page 181).

To illustrate the basics of portfolio diversification, consider an investor who forms a portfolio, consisting of only two stocks, by investing $4,000 in one stock and $6,000 in a second stock. Suppose that the results after 1 year are as listed here. (We've previously defined return on investment. See Applications in Finance: Return on Investment on page 52.)

One-Year Results

Stock	Initial Investment ($)	Value of Investment After One Year ($)	Rate of Return on Investment
1	4,000	5,000	$R_1 = .25$ (25%)
2	6,000	5,400	$R_2 = -.10$ (-10%)
Total	10,000	10,400	$R_p = .04$ (4%)

Another way of calculating the portfolio return R_p is to compute the weighted average of the individual stock returns R_1 and R_2, where the weights w_1 and w_2 are the proportions of the initial $10,000 invested in stocks 1 and 2, respectively. In this illustration, $w_1 = .4$ and $w_2 = .6$. (Note that w_1 and w_2 must always sum to 1 because the two stocks constitute the entire portfolio.) The weighted average of the two returns is

$$\begin{aligned} R_p &= w_1R_1 + w_2R_2 \\ &= (.4)(.25) + (.6)(-.10) = .04 \end{aligned}$$

This is how portfolio returns are calculated. However, when the initial investments are made, the investor does not know what the returns will be. In fact, the returns are random variables. We are interested in determining the expected value and variance of the portfolio. The formulas in the box were derived from the laws of expected value and variance introduced in the two previous sections.

Mean and Variance of a Portfolio of Two Stocks

$$\begin{aligned} E(R_p) &= w_1E(R_1) + w_2E(R_2) \\ V(R_p) &= w_1^2V(R_1) + w_2^2V(R_2) + 2w_1w_2\text{COV}(R_1,R_2) \\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2 \end{aligned}$$

where w_1 and w_2 are the proportions or weights of investments 1 and 2, $E(R_1)$ and $E(R_2)$ are their expected values, σ_1 and σ_2 are their standard deviations, $\text{COV}(R_1,R_2)$ is the covariance, and ρ is the coefficient of correlation.

(Recall that $\rho = \dfrac{\text{COV}(R_1,R_2)}{\sigma_1\sigma_2}$, which means that $\text{COV}(R_1,R_2) = \rho\sigma_1\,\sigma_2$.)

EXAMPLE 7.8 Describing the Population of the Returns on a Portfolio

An investor has decided to form a portfolio by putting 25% of his money into McDonald's stock and 75% into Cisco Systems stock. The investor assumes that the expected returns will be 8% and 15%, respectively, and that the standard deviations will be 12% and 22%, respectively.

a. Find the expected return on the portfolio.
b. Compute the standard deviation of the returns on the portfolio assuming that
   i. the two stocks' returns are perfectly positively correlated.
   ii. the coefficient of correlation is .5.
   iii. the two stocks' returns are uncorrelated.

SOLUTION

a. The expected values of the two stocks are

$$E(R_1) = .08 \quad \text{and} \quad E(R_2) = .15$$

The weights are $w_1 = .25$ and $w_2 = .75$.
Thus,

$$E(R_p) = w_1E(R_1) + w_2E(R_2) = .25(.08) + .75(.15) = .1325$$

b. The standard deviations are

$$\sigma_1 = .12 \quad \text{and} \quad \sigma_2 = .22$$

Thus,

$$\begin{aligned} V(R_p) &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2 \\ &= (.25^2)(.12^2) + (.75^2)(.22^2) + 2(.25)(.75)\,\rho\,(.12)(.22) \\ &= .0281 + .0099\rho \end{aligned}$$

When $\rho = 1$

$$V(R_p) = .0281 + .0099(1) = .0380$$

$$\text{Standard deviation} = \sqrt{V(R_p)} = \sqrt{.0380} = .1949$$

When $\rho = .5$

$$V(R_p) = .0281 + .0099(.5) = .0331$$

$$\text{Standard deviation} = \sqrt{V(R_p)} = \sqrt{.0331} = .1819$$

When $\rho = 0$

$$V(R_p) = .0281 + .0099(0) = .0281$$

$$\text{Standard deviation} = \sqrt{V(R_p)} = \sqrt{.0281} = .1676$$

Notice that the variance and standard deviation of the portfolio returns decrease as the coefficient of correlation decreases.

Portfolio Diversification in Practice

The formulas introduced in this section require that we know the expected values, variances, and covariance (or coefficient of correlation) of the investments we're interested in. The question arises, how do we determine these parameters? (Incidentally, this question is rarely addressed in finance textbooks!) The most common procedure is to estimate the parameters from historical data, using sample statistics.

Portfolios with More Than Two Stocks

We can extend the formulas that describe the mean and variance of the returns of a portfolio of two stocks to a portfolio of any number of stocks.

Mean and Variance of a Portfolio of k Stocks

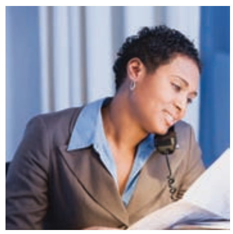
$$E(R_p) = \sum_{i=1}^{k} w_i E(R_i)$$

$$V(R_p) = \sum_{i=1}^{k} w_i^2 \sigma_i^2 + 2\sum_{i=1}^{k}\sum_{j=i+1}^{k} w_i w_j \mathrm{COV}(R_i, R_j)$$

Where R_i is the return of the ith stock, w_i is the proportion of the portfolio invested in stock i, and k is the number of stocks in the portfolio.

When k is greater than 2, the calculations can be tedious and time consuming. For example, when $k = 3$, we need to know the values of the three weights, three expected values, three variances, and three covariances. When $k = 4$, there are four expected values, four variances, and six covariances. [The number of covariances required in general is $k(k - 1)/2$.] To assist you, we have created an Excel worksheet to perform the computations when $k = 2$, 3, or 4. To demonstrate, we'll return to the problem described in this chapter's introduction.

Investing to Maximize Returns and Minimize Risk: Solution



© Terry Vine/Blend Images/ Jupiterimages

Because of the large number of calculations, we will solve this problem using only Excel. From the file, we compute the means of each stock's returns.

Excel Means

	A	B	C	D
1	0.00881	0.00562	0.01253	0.02834

Next we compute the variance–covariance matrix. (The commands are the same as those described in Chapter 4: Simply include all the columns of the returns of the investments you wish to include in the portfolio.)

Excel Variance-Covariance Matrix

	A	B	C	D	E
1		Coca Cola	Disney	Barrick	Amazon
2	Coca Cola	0.00235			
3	Disney	0.00141	0.00434		
4	Barrick	0.00184	–0.00058	0.01174	
5	Amazon	0.00167	0.00182	–0.00170	0.02020

Notice that the variances of the returns are listed on the diagonal. Thus, for example, the variance of the 60 monthly returns of Barrick Gold is .01174. The covariances appear below the diagonal. The covariance between the returns of Coca Cola and Disney is .00141.

The means and the variance–covariance matrix are copied to the spreadsheet using the commands described here. The weights are typed producing the accompanying output.

Excel Worksheet: Portfolio Diversification—Plan 1

	A	B	C	D	E	F
1	**Portfolio of 4 Stocks**					
2			**Coca Cola**	**Disney**	**Barrick**	**Amazon**
3	**Variance-Covariance Matrix**	**Coca Cola**	0.00235			
4		**Disney**	0.00141	0.00434		
5		**Barrick**	0.00184	−0.00058	0.01174	
6		**Amazon**	0.00167	0.00182	−0.00170	0.02020
7						
8	**Expected Returns**		0.00881	0.00562	0.01253	0.02834
9						
10	**Weights**		0.25000	0.25000	0.25000	0.25000
11						
12	**Portfolio Return**					
13	**Expected Value**	**0.01382**				
14	**Variance**	**0.00297**				
15	**Standard Deviation**	**0.05452**				

The expected return on the portfolio is .01382, and the variance is .00297.

INSTRUCTIONS

1. Open the file containing the returns. In this example, open file Ch7:\ Xm07-00
2. Compute the means of the columns containing the returns of the stocks in the portfolio.
3. Using the commands described in Chapter 4 (page 137) compute the variance–covariance matrix.
4. Open the **Portfolio Diversification** workbook. Use the tab to select the **4 Stocks** worksheet. Do not change any cells that appear in bold print. Do not save any worksheets.
5. Copy the means into cells C8 to F8. (Use **Copy**, **Paste Special** with **Values and number formats**.)
6. Copy the variance–covariance matrix (including row and column labels) into columns B, C, D, E, and F.
7. Type the weights into cells C10 to F10

The mean, variance, and standard deviation of the portfolio will be printed. Use similar commands for 2 stock and 3 stock portfolios.

The results for plan 2 are

	A	B
12	**Portfolio Return**	
13	**Expected Value**	**0.01710**
14	**Variance**	**0.00460**
15	**Standard Deviation**	**0.06783**

Plan 3

	A	B
12	**Portfolio Return**	
13	**Expected Value**	**0.01028**
14	**Variance**	**0.00256**
15	**Standard Deviation**	**0.05059**

Plan 3 has the smallest expected value and the smallest variance. Plan 2 has the largest expected value and the largest variance. Plan 1's expected value and variance are in the middle. If the investor is like most investors, she would select Plan 3 because of its lower risk. Other, more daring investors may choose plan 2 to take advantage of its higher expected value.

In this example, we showed how to compute the expected return, variance, and standard deviation from a sample of returns on the investments for any combination of weights. (We illustrated the process with three sets of weights.) It is possible to determine the "optimal" weights that minimize risk for a given expected value or maximize expected return for a given standard deviation. This is an extremely important function of financial analysts and investment advisors. Solutions can be determined using a management-science technique called *linear programming*, a subject taught by most schools of business and faculties of management.

EXERCISES

7.61 Describe what happens to the expected value and standard deviation of the portfolio returns when the coefficient of correlation decreases.

7.62 A portfolio is composed of two stocks. The proportion of each stock, their expected values, and standard deviations are listed next.

Stock	1	2
Proportion of portfolio	.30	.70
Mean	.12	.25
Standard deviation	.02	.15

For each of the following coefficients of correlation calculate the expected value and standard deviation of the portfolio.

a. $\rho = .5$
b. $\rho = .2$
c. $\rho = 0$

7.63 An investor is given the following information about the returns on two stocks.

Stock	1	2
Mean	.09	.13
Standard deviation	.15	.21

a. If he is most interested in maximizing his returns, which stock should he choose?
b. If he is most interested in minimizing his risk, which stock should he choose?

7.64 Refer to Exercise 7.63. Compute the expected value and standard deviation of the portfolio composed of 60% stock 1 and 40% stock 2. The coefficient of correlation is .4.

7.65 Refer to Exercise 7.63. Compute the expected value and standard deviation of the portfolio composed of 30% stock 1 and 70% stock 2.

The following exercises require the use of a computer.

Xr07-66 *The monthly returns for the following stocks on the New York Stock Exchange were recorded.*

AT&T, Aetna, Cigna, Coca-Cola, Disney, Ford, and McDonald's

The next seven exercises are based on this set of data.

7.66 a. Calculate the mean and variance of the monthly return for each stock.
b. Determine the variance–covariance matrix.

7.67 Select the two stocks with the largest means and construct a portfolio consisting of equal amounts of both. Determine the expected value and standard deviation of the portfolio.

7.68 Select the two stocks with the smallest variances and construct a portfolio consisting of equal amounts of both. Determine the expected value and standard deviation of the portfolio.

7.69 Describe the results of Exercises 7.66 to 7.68.

7.70 An investor wants to develop a portfolio composed of shares of AT&T, Coca-Cola, Ford, and Disney. Calculate the expected value and standard deviation of the returns for a portfolio with equal proportions of all three stocks.

7.71 Suppose you want a portfolio composed of AT&T, Cigna, Disney, and Ford. Find the expected value and standard deviation of the returns for the following portfolio.

AT&T	30%
Cigna	20%
Disney	40%
Ford	10%

7.72 Repeat Exercise 7.71 using the following proportions. Compare your results with those of Exercise 7.71.

AT&T	30%
Cigna	10%
Disney	40%
Ford	20%

The following seven exercises are directed at Canadian students.

Xr07-73 *The monthly returns for the following stocks on the Toronto Stock Exchange were recorded:*

Barrick Gold, Bell Canada Enterprises (BCE),Bank of Montreal (BMO), Enbridge, Fortis, Methanex, Research in Motion, Telus, and Trans Canada Pipeline

The next seven exercises are based on this set of data.

7.73 a. Calculate the mean and variance of the monthly return for each stock.
b. Determine the correlation matrix.

7.74 Select the two stocks with the largest means and construct a portfolio consisting of equal amounts of both. Determine the expected value and standard deviation of the portfolio.

7.75 Select the two stocks with the smallest variances and construct a portfolio consisting of equal amounts of both. Determine the expected value and standard deviation of the portfolio.

7.76 Describe the results of Exercises 7.73 to 7.75.

7.77 An investor wants to develop a portfolio composed of shares of Bank of Montreal, Enbridge, and Fortis. Calculate the expected value and standard deviation of the returns for a portfolio with the following proportions.

Bank of Montreal	20%
Enbridge	30%
Fortis	50%

7.78 Suppose you want a portfolio composed of Barrick Gold, Bell Canada Enterprises, Telus, and Trans-Canada Ltd. Find the expected value and standard deviation of the returns for the following portfolio.

Barrick Gold	50%
Bell Canada Enterprises	25%
Telus	15%
TransCanada	10%

7.79 Repeat Exercise 7.78 using the following proportions. Compare your results with those of Exercise 7.78.

Barrick Gold	20%
Bell Canada Enterprises	40%
Telus	20%
TransCanada	20%

Xr07-80 *The monthly returns for the following stocks on the NASDAQ Stock Exchange were recorded:*

Amazon, Amgen, Apple, Cisco Systems, Google, Intel, Microsoft, Oracle, and Research in Motion

The next four exercises are based on this set of data.

7.80 a. Calculate the mean and variance of the monthly return for each stock.
b. Determine which four stocks you would include in your portfolio if you wanted a large expected value.
c. Determine which four stocks you would include in your portfolio if you wanted a small variance.

7.81 Suppose you want a portfolio composed of Cisco Systems, Intel, Microsoft, and Research in Motion. Find the expected value and standard deviation of the returns for the following portfolio.

Cisco Systems	30%
Intel	15%
Microsoft	25%
Research in Motion	30%

7.82 An investor wants to acquire a portfolio composed of Cisco Systems, Intel, Microsoft, and Research in Motion. Moreover, he wants the expected value to be at least 1%. Try several sets of proportions (remember, they must add to 1.0) to see if you can find the portfolio with the smallest variance.

7.83 Refer to Exercise 7.81.
a. Compute the expected value and variance of the portfolio described next.

Cisco Systems	26.59%
Intel	2.49%
Microsoft	54.74%
Research in Motion	16.19%

b. Can you do better? In other words, can you find a portfolio whose expected value is greater than or equal to 1% and whose variance is less than the one you calculated in part (a)? (*Hint:* Don't spend too much time at this. You won't be able to do better.)
c. If you want to learn how we produced the portfolio above, take a course that teaches linear and nonlinear programming.

7.4 BINOMIAL DISTRIBUTION

Now that we've introduced probability distributions in general, we need to introduce several specific probability distributions. In this section, we present the *binomial distribution*.

The binomial distribution is the result of a *binomial experiment*, which has the following properties.

Binomial Experiment

1. The **binomial experiment** consists of a fixed number of trials. We represent the number of trials by n.
2. Each trial has two possible outcomes. We label one outcome a *success*, and the other a *failure*.
3. The probability of success is p. The probability of failure is $1 - p$.
4. The trials are independent, which means that the outcome of one trial does not affect the outcomes of any other trials.

If properties 2, 3, and 4 are satisfied, we say that each trial is a **Bernoulli process**. Adding property 1 yields the binomial experiment. The random variable of a binomial experiment is defined as the number of successes in the n trials. It is called the **binomial random variable**. Here are several examples of binomial experiments.

1. Flip a coin 10 times. The two outcomes per trial are heads and tails. The terms *success* and *failure* are arbitrary. We can label either outcome success. However, generally, we call success anything we're looking for. For example, if we were betting on heads, we would label heads a success. If the coin is fair, the probability of heads is 50%. Thus, $p = .5$. Finally, we can see that the trials are independent because the outcome of one coin flip cannot possibly affect the outcomes of other flips.

2. Draw five cards out of a shuffled deck. We can label as success whatever card we seek. For example, if we wish to know the probability of receiving five clubs, a club is labeled a success. On the first draw, the probability of a club is 13/52 = .25. However, if we draw a second card without replacing the first card and shuffling, the trials are not independent. To see why, suppose that the first draw is a club. If we draw again without replacement the probability of drawing a second club is 12/51, which is not .25. In this experiment, the trials are *not* independent.* Hence, this is not a binomial experiment. However, if we replace the card and shuffle before drawing again, the experiment is binomial. Note that in most card games, we do not replace the card, and as a result the experiment is not binomial.

3. A political survey asks 1,500 voters who they intend to vote for in an approaching election. In most elections in the United States, there are only two candidates, the Republican and Democratic nominees. Thus, we have two outcomes per trial. The trials are independent because the choice of one voter does not affect the choice of other voters. In Canada, and in other countries with parliamentary systems of government, there are usually several candidates in the race. However, we can label a vote for our favored candidate (or the party that is paying us to do the survey) a success and all the others are failures.

As you will discover, the third example is a very common application of statistical inference. The actual value of p is unknown, and the job of the statistics practitioner is to estimate its value. By understanding the probability distribution that uses p, we will be able to develop the statistical tools to estimate p.

*The hypergeometric distribution described in Keller's website Appendix of the same name is used to calculate probabilities in such cases.

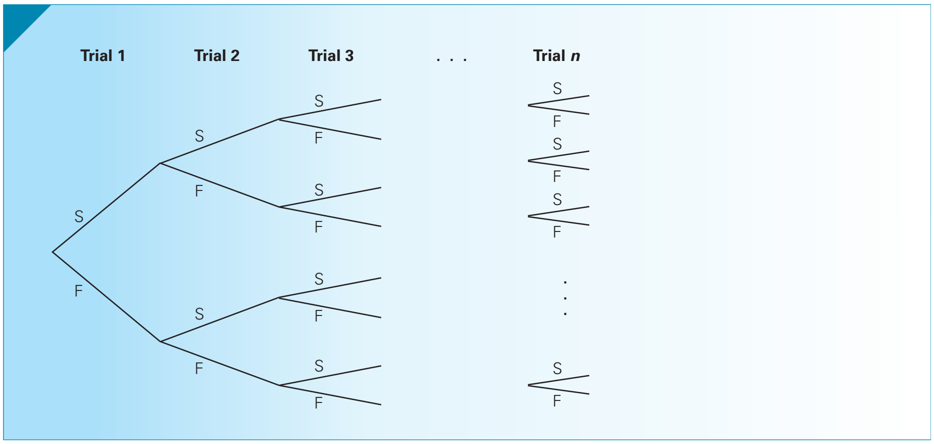
Binomial Random Variable

The binomial random variable is the number of successes in the experiment's n trials. It can take on values 0, 1, 2, . . . , n. Thus, the random variable is discrete. To proceed, we must be capable of calculating the probability associated with each value.

Using a probability tree, we draw a series of branches as depicted in Figure 7.2. The stages represent the outcomes for each of the n trials. At each stage, there are two branches representing success and failure. To calculate the probability that there are X successes in n trials, we note that for each success in the sequence we must multiply by p. And if there are X successes, there must be $n - X$ failures. For each failure in the sequence, we multiply by $1 - p$. Thus, the probability for each sequence of branches that represent x successes and $n - x$ failures has probability

$$p^x(1 - p)^{n-x}$$

FIGURE **7.2** **Probability Tree for a Binomial Experiment**



There are a number of branches that yield x successes and $n - x$ failures. For example, there are two ways to produce exactly one success and one failure in two trials: SF and FS. To count the number of branch sequences that produce x successes and $n - x$ failures, we use the combinatorial formula

$$C_x^n = \frac{n!}{x!(n - x)!}$$

where $n! = n(n - 1)\,(n - 2) \ldots (2)(1)$. For example, $3! = 3(2)(1) = 6$. Incidentally, although it may not appear to be logical $0! = 1$.

Pulling together the two components of the probability distribution yields the following.

Binomial Probability Distribution

The probability of x successes in a binomial experiment with n trials and probability of success $= p$ is

$$P(x) = \frac{n!}{x!(n - x)!}p^x(1 - p)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n$$

EXAMPLE 7.9 Pat Statsdud and the Statistics Quiz

Pat Statsdud is a student taking a statistics course. Unfortunately, Pat is not a good student. Pat does not read the textbook before class, does not do homework, and regularly misses class. Pat intends to rely on luck to pass the next quiz. The quiz consists of 10 multiple-choice questions. Each question has five possible answers, only one of which is correct. Pat plans to guess the answer to each question.

a. What is the probability that Pat gets no answers correct?

b. What is the probability that Pat gets two answers correct?

SOLUTION

The experiment consists of 10 identical trials, each with two possible outcomes and where success is defined as a correct answer. Because Pat intends to guess, the probability of success is 1/5 or .2. Finally, the trials are independent because the outcome of any of the questions does not affect the outcomes of any other questions. These four properties tell us that the experiment is binomial with $n = 10$ and $p = .2$.

a. From

$$P(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

we produce the probability of no successes by letting $n = 10$, $p = .2$, and $x = 0$. Hence,

$$P(0) = \frac{10!}{0!(10-0)!}(.2)^0(1-.2)^{10-0}$$

The combinatorial part of the formula is $\frac{10!}{0!10!}$, which is 1. This is the number of ways to get 0 correct and 10 incorrect. Obviously, there is only one way to produce $X = 0$. And because $(.2)^0 = 1$,

$$P(X = 0) = 1(1)(.8)^{10} = .1074$$

b. The probability of two correct answers is computed similarly by substituting $n = 10$, $p = .2$, and $x = 2$:

$$P(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$$P(0) = \frac{10!}{2!(10-2)!}(.2)^2(1-.2)^{10-2}$$

$$= \frac{(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)}{(2)(1)(8)(7)(6)(5)(4)(3)(2)(1)}(.04)(.1678)$$

$$= 45(.006712)$$

$$= .3020$$

In this calculation, we discovered that there are 45 ways to get exactly two correct and eight incorrect answers, and that each such outcome has probability .006712. Multiplying the two numbers produces a probability of .3020.

Cumulative Probability

The formula of the binomial distribution allows us to determine the probability that X equals individual values. In Example 7.9, the values of interest were 0 and 2. There are many circumstances where we wish to find the probability that a random variable is less than or equal to a value; that is, we want to determine $P(X \leq x)$, where x is that value. Such a probability is called a **cumulative probability**.

EXAMPLE 7.10

Will Pat Fail the Quiz?

Find the probability that Pat fails the quiz. A mark is considered a failure if it is less than 50%.

SOLUTION

In this quiz, a mark of less than 5 is a failure. Because the marks must be integers, a mark of 4 or less is a failure. We wish to determine $P(X \leq 4)$. So,

$$P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$$

From Example 7.9, we know $P(0) = .1074$ and $P(2) = .3020$. Using the binomial formula, we find $P(1) = .2684$, $P(3) = .2013$, and $P(4) = .0881$. Thus

$$P(X \leq 4) = .1074 + .2684 + .3020 + .2013 + .0881 = .9672$$

There is a 96.72% probability that Pat will fail the quiz by guessing the answer for each question.

Binomial Table

There is another way to determine binomial probabilities. Table 1 in Appendix B provides cumulative binomial probabilities for selected values of n and p. We can use this table to answer the question in Example 7.10, where we need $P(X \leq 4)$. Refer to Table 1, find $n = 10$, and in that table find $p = .20$. The values in that column are $P(X \leq x)$ for $x = 0, 1, 2, \ldots, 10$, which are shown in Table 7.2.

TABLE **7.2** Cumulative Binomial Probabilities with $n = 10$ and $p = .2$

x	$P(X \leq x)$
0	.1074
1	.3758
2	.6778
3	.8791
4	.9672
5	.9936
6	.9991
7	.9999
8	1.000
9	1.000
10	1.000

The first cumulative probability is $P(X \le 0)$, which is $P(0) = .1074$. The probability we need for Example 7.10 is $P(X \le 4) = .9672$, which is the same value we obtained manually.

We can use the table and the complement rule to determine probabilities of the type $P(X \ge x)$. For example, to find the probability that Pat will pass the quiz, we note that

$$P(X \le 4) + P(X \ge 5) = 1$$

Thus,

$$P(X \ge 5) = 1 - P(X \le 4) = 1 - .9672 = .0328$$

Using Table 1 to Find the Binomial Probability $P(X \ge x)$

$$P(X \ge x) = 1 - P(X \le [x - 1])$$

The table is also useful in determining the probability of an individual value of X. For example, to find the probability that Pat will get exactly two right answers we note that

$$P(X \le 2) = P(0) + P(1) + P(2)$$

and

$$P(X \le 1) = P(0) + P(1)$$

The difference between these two cumulative probabilities is $p(2)$. Thus,

$$P(2) = P(X \le 2) - P(X \le 1) = .6778 - .3758 = .3020$$

Using Table 1 to Find the Binomial Probability $P(X = x)$

$$P(x) = P(X \le x) - P(X \le [x - 1])$$

Using the Computer

EXCEL

INSTRUCTIONS

Type the following into any empty cell:

=**BINOMDIST**([x], [n], [p], [True] or [False])

Typing “True” calculates a cumulative probability and typing “False” computes the probability of an individual value of X. For Example 7.9(a), type

=**BINOMDIST**(0, 10, .2, False)

For Example 7.10, enter

=**BINOMDIST**(4, 10, .2, True)

MINITAB

INSTRUCTIONS

This is the first of seven probability distributions for which we provide instructions. All work in the same way. Click **Calc**, **Probability Distributions**, and the specific distribution whose probability you wish to compute. In this case, select **Binomial . . . .** Check either **Probability** or **Cumulative probability**. If you wish to make a probability statement about one value of x, specify **Input constant** and type the value of x.

If you wish to make probability statements about several values of x from the same binomial distribution, type the values of x into a column before checking **Calc**. Choose **Input column** and type the name of the column. Finally, enter the components of the distribution. For the binomial, enter the **Number of trials** n and the **Event Probability** p.

For the other six distributions, we list the distribution (here it is **Binomial**) and the components only (for this distribution it is n and p).

Mean and Variance of a Binomial Distribution

Statisticians have developed general formulas for the mean, variance, and standard deviation of a binomial random variable. They are

$$\mu = np$$
$$\sigma^2 = np(1 - p)$$
$$\sigma = \sqrt{np(1 - p)}$$

EXAMPLE 7.11

Pat Statsdud Has Been Cloned!

Suppose that a professor has a class full of students like Pat (a nightmare!). What is the mean mark? What is the standard deviation?

SOLUTION

The mean mark for a class of Pat Statsduds is

$$\mu = np = 10(.2) = 2$$

The standard deviation is

$$\sigma = \sqrt{np(1 - p)} = \sqrt{10(.2)(1 - .2)} = 1.26$$

EXERCISES

7.84 Given a binomial random variable with $n = 10$ and $p = .3$, use the formula to find the following probabilities.
a. $P(X = 3)$
b. $P(X = 5)$
c. $P(X = 8)$

7.85 Repeat Exercise 7.84 using Table 1 in Appendix B.

7.86 Repeat Exercise 7.84 using Excel or Minitab.

7.87 Given a binomial random variable with $n = 6$ and $p = .2$, use the formula to find the following probabilities.
a. $P(X = 2)$
b. $P(X = 3)$
c. $P(X = 5)$

7.88 Repeat Exercise 7.87 using Table 1 in Appendix B.

7.89 Repeat Exercise 7.87 using Excel or Minitab.

7.90 Suppose X is a binomial random variable with $n = 25$ and $p = .7$. Use Table 1 to find the following.
a. $P(X = 18)$
b. $P(X = 15)$
c. $P(X \leq 20)$
d. $P(X \geq 16)$

7.91 Repeat Exercise 7.90 using Excel or Minitab.

7.92 A sign on the gas pumps of a chain of gasoline stations encourages customers to have their oil checked with the claim that one out of four cars needs to have oil added. If this is true, what is the probability of the following events?
a. One out of the next four cars needs oil
b. Two out of the next eight cars need oil
c. Three out of the next 12 cars need oil

7.93 The leading brand of dishwasher detergent has a 30% market share. A sample of 25 dishwasher detergent customers was taken. What is the probability that 10 or fewer customers chose the leading brand?

7.94 A certain type of tomato seed germinates 90% of the time. A backyard farmer planted 25 seeds.
a. What is the probability that exactly 20 germinate?
b. What is the probability that 20 or more germinate?
c. What is the probability that 24 or fewer germinate?
d. What is the expected number of seeds that germinate?

7.95 According to the American Academy of Cosmetic Dentistry, 75% of adults believe that an unattractive smile hurts career success. Suppose that 25 adults are randomly selected. What is the probability that 15 or more of them would agree with the claim?

7.96 A student majoring in accounting is trying to decide on the number of firms to which he should apply. Given his work experience and grades, he can expect to receive a job offer from 70% of the firms to which he applies. The student decides to apply to only four firms. What is the probability that he receives no job offers?

7.97 In the United States, voters who are neither Democrat nor Republican are called Independents. It is believed that 10% of all voters are Independents. A survey asked 25 people to identify themselves as Democrat, Republican, or Independent.
a. What is the probability that none of the people are Independent?
b. What is the probability that fewer than five people are Independent?
c. What is the probability that more than two people are Independent?

7.98 Most dial-up Internet service providers (ISPs) attempt to provide a large enough service so that customers seldom encounter a busy signal. Suppose that the customers of one ISP encounter busy signals 8% of the time. During the week, a customer of this ISP called 25 times. What is the probability that she did not encounter any busy signals?

7.99 Major software manufacturers offer a help line that allows customers to call and receive assistance in solving their problems. However, because of the volume of calls, customers frequently are put on hold. One software manufacturer claims that only 20% of callers are put on hold. Suppose that 100 customers call. What is the probability that more than 25 of them are put on hold?

7.100 A statistics practitioner working for major league baseball determined the probability that the hitter will be out on ground balls is .75. In a game where there are 20 ground balls, find the probability that all of them were outs.

The following exercises are best solved with a computer.

7.101 The probability of winning a game of craps (a dice-throwing game played in casinos) is 244/495.
a. What is the probability of winning 5 or more times in 10 games?
b. What is the expected number of wins in 100 games?

7.102 In the game of blackjack as played in casinos in Las Vegas, Atlantic City, and Niagara Falls, as well as in many other cities, the dealer has the advantage. Most players do not play very well. As a result, the probability that the average player wins a hand is about 45%. Find the probability that an average player wins.
a. Twice in 5 hands
b. Ten or more times in 25 hands

7.103 Several books teach blackjack players the "basic strategy," which increases the probability of winning any hand to 50%. Repeat Exercise 7.102, assuming the player plays the basic strategy.

7.104 The best way of winning at blackjack is to "case the deck," which involves counting 10s, non-10s, and aces. For card counters, the probability of winning a hand may increase to 52%. Repeat Exercise 7.102 for a card counter.

7.105 In the game of roulette, a steel ball is rolled onto a wheel that contains 18 red, 18 black, and 2 green slots. If the ball is rolled 25 times, find the probabilities of the following events.
a. The ball falls into the green slots two or more times.
b. The ball does not fall into any green slots.
c. The ball falls into black slots 15 or more times.
d. The ball falls into red slots 10 or fewer times.

7.106 According to a Gallup Poll conducted March 5–7, 2001, 52% of American adults think that protecting the environment should be given priority over developing U.S. energy supplies. Thirty-six percent think that developing energy supplies is more important, and 6% believe the two are equally important. The rest had no opinion. Suppose that a sample of 100 American adults is quizzed on the subject. What is the probability of the following events?

a. Fifty or more think that protecting the environment should be given priority.
b. Thirty or fewer think that developing energy supplies is more important.
c. Five or fewer have no opinion.

7.107 In a *Bon Appetit* poll, 38% of people said that chocolate was their favorite flavor of ice cream. A sample of 20 people was asked to name their favorite flavor of ice cream. What is the probability that half or more of them prefer chocolate?

7.108 The statistics practitioner in Exercise 7.100 also determined that if a batter hits a line drive, the probability of an out is .23. Determine the following probabilities.

a. In a game with 10 line drives, at least 5 are outs.
b. In a game with 25 line drives, there are 5 outs or less.

7.109 According to the last census, 45% of working women held full-time jobs in 2002. If a random sample of 50 working women is drawn, what is the probability that 19 or more hold full-time jobs?

7.5 Poisson Distribution

Another useful discrete probability distribution is the **Poisson distribution**, named after its French creator. Like the binomial random variable, the **Poisson random variable** is the number of occurrences of events, which we'll continue to call *successes*. The difference between the two random variables is that a binomial random variable is the number of successes in a set number of trials, whereas a Poisson random variable is the number of successes in an interval of time or specific region of space. Here are several examples of Poisson random variables.

1. The number of cars arriving at a service station in 1 hour. (The interval of time is 1 hour.)
2. The number of flaws in a bolt of cloth. (The specific region is a bolt of cloth.)
3. The number of accidents in 1 day on a particular stretch of highway. (The interval is defined by both time, 1 day, and space, the particular stretch of highway.)

The Poisson experiment is described in the box.

Poisson Experiment

A **Poisson experiment** is characterized by the following properties:

1. The number of successes that occur in any interval is independent of the number of successes that occur in any other interval.
2. The probability of a success in an interval is the same for all equal-size intervals.
3. The probability of a success in an interval is proportional to the size of the interval.
4. The probability of more than one success in an interval approaches 0 as the interval becomes smaller.

Poisson Random Variable

The **Poisson random variable** is the number of successes that occur in a period of time or an interval of space in a Poisson experiment.

There are several ways to derive the probability distribution of a Poisson random variable. However, all are beyond the mathematical level of this book. We simply provide the formula and illustrate how it is used.

Poisson Probability Distribution

The probability that a Poisson random variable assumes a value of x in a specific interval is

$$P(x) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{for } x = 0,1,2,\ldots$$

where μ is the mean number of successes in the interval or region and e is the base of the natural logarithm (approximately 2.71828). Incidentally, the variance of a Poisson random variable is equal to its mean; that is, $\sigma^2 = \mu$.

EXAMPLE 7.12 Probability of the Number of Typographical Errors in Textbooks

A statistics instructor has observed that the number of typographical errors in new editions of textbooks varies considerably from book to book. After some analysis, he concludes that the number of errors is Poisson distributed with a mean of 1.5 per 100 pages. The instructor randomly selects 100 pages of a new book. What is the probability that there are no typographical errors?

SOLUTION

We want to determine the probability that a Poisson random variable with a mean of 1.5 is equal to 0. Using the formula

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

and substituting $x = 0$ and $\mu = 1.5$, we get

$$P(0) = \frac{e^{-1.5}1.5^0}{0!} = \frac{(2.71828)^{-1.5}(1)}{1} = .2231$$

The probability that in the 100 pages selected there are no errors is .2231.

Notice that in Example 7.12 we wanted to find the probability of 0 typographical errors in 100 pages given a mean of 1.5 typos in 100 pages. The next example illustrates how we calculate the probability of events where the intervals or regions do not match.

EXAMPLE 7.13

Probability of the Number of Typographical Errors in 400 Pages

Refer to Example 7.12. Suppose that the instructor has just received a copy of a new statistics book. He notices that there are 400 pages.

a. What is the probability that there are no typos?
b. What is the probability that there are five or fewer typos?

SOLUTION

The specific region that we're interested in is 400 pages. To calculate Poisson probabilities associated with this region, we must determine the mean number of typos per 400 pages. Because the mean is specified as 1.5 per 100 pages, we multiply this figure by 4 to convert to 400 pages. Thus, $\mu = 6$ typos per 400 pages.

a. The probability of no typos is

$$P(0) = \frac{e^{-6}6^0}{0!} = \frac{(2.71828)^{-6}(1)}{1} = .002479$$

b. We want to determine the probability that a Poisson random variable with a mean of 6 is 5 or less; that is, we want to calculate

$$P(X \leq 5) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$$

To produce this probability, we need to compute the six probabilities in the summation.

$$P(0) = .002479$$

$$P(1) = \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-6}6^1}{1!} = \frac{(2.71828)^{-6}(6)}{1} = .01487$$

$$P(2) = \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-6}6^2}{2!} = \frac{(2.71828)^{-6}(36)}{2} = .04462$$

$$P(3) = \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-6}6^3}{3!} = \frac{(2.71828)^{-6}(216)}{6} = .08924$$

$$P(4) = \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-6}6^4}{4!} = \frac{(2.71828)^{-6}(1296)}{24} = .1339$$

$$P(5) = \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-6}6^5}{5!} = \frac{(2.71828)^{-6}(7776)}{120} = .1606$$

Thus,

$$P(X \leq 5) = .002479 + .01487 + .04462 + .08924 + .1339 + .1606$$
$$= .4457$$

The probability of observing 5 or fewer typos in this book is .4457.

Poisson Table

As was the case with the binomial distribution, a table is available that makes it easier to compute Poisson probabilities of individual values of x as well as cumulative and related probabilities.

Table 2 in Appendix B provides cumulative Poisson probabilities for selected values of μ. This table makes it easy to find cumulative probabilities like those in Example 7.13, part (b), where we found $P(X \leq 5)$.

To do so, find $\mu = 6$ in Table 2. The values in that column are $P(X \leq x)$ for $x = 0, 1, 2, \ldots, 18$ which are shown in Table 7.3.

TABLE **7.3** Cumulative Poisson Probabilities for $\mu = 6$

x	$P(X \leq x)$
0	.0025
1	.0174
2	.0620
3	.1512
4	.2851
5	.4457
6	.6063
7	.7440
8	.8472
9	.9161
10	.9574
11	.9799
12	.9912
13	.9964
14	.9986
15	.9995
16	.9998
17	.9999
18	1.0000

Theoretically, a Poisson random variable has no upper limit. The table provides cumulative probabilities until the sum is 1.0000 (using four decimal places).

The first cumulative probability is $P(X \leq 0)$, which is $P(0) = .0025$. The probability we need for Example 7.13, part (b), is $P(X \leq 5) = .4457$, which is the same value we obtained manually.

Like Table 1 for binomial probabilities, Table 2 can be used to determine probabilities of the type $P(X \geq x)$. For example, to find the probability that in Example 7.13 there are 6 or more typos, we note that $P(X \leq 5) + P(X \geq 6) = 1$. Thus,

$$P(X \geq 6) = 1 - P(X \leq 5) = 1 - .4457 = .5543$$

Using Table 2 to Find the Poisson Probability $P(X \geq x)$

$$P(X \geq x) = 1 - P(X \leq [x - 1])$$

We can also use the table to determine the probability of one individual value of X. For example, to find the probability that the book contains exactly 10 typos, we note that

$$P(X \leq 10) = P(0) + P(1) + \cdots + P(9) + P(10)$$

and

$$P(X \leq 9) = P(0) + P(1) + \cdots + P(9)$$

The difference between these two cumulative probabilities is $P(10)$. Thus,

$$P(10) = P(X \leq 10) - P(X \leq 9) = .9574 - .9161 = .0413$$

Using Table 2 to Find the Poisson Probability $P(X = x)$

$$P(x) = P(X \leq x) - P(X \leq [x - 1])$$

Using the Computer

EXCEL

INSTRUCTIONS

Type the following into any empty cell:

= **POISSON**([x], [μ], [True] or [False])

We calculate the probability in Example 7.12 by typing

= **POISSON**(0, 1.5, False)

For Example 7.13, we type

= **POISSON**(5, 6, True)

MINITAB

INSTRUCTIONS

Click **Calc**, **Probability Distributions**, and **Poisson . . .** and type the mean.

EXERCISES

7.110 Given a Poisson random variable with $\mu = 2$, use the formula to find the following probabilities.
a. $P(X = 0)$
b. $P(X = 3)$
c. $P(X = 5)$

7.111 Given that X is a Poisson random variable with $\mu = .5$, use the formula to determine the following probabilities.
a. $P(X = 0)$
b. $P(X = 1)$
c. $P(X = 2)$

7.112 The number of accidents that occur at a busy intersection is Poisson distributed with a mean of 3.5 per week. Find the probability of the following events.
a. No accidents in one week
b. Five or more accidents in one week
c. One accident today

7.113 Snowfalls occur randomly and independently over the course of winter in a Minnesota city. The average is one snowfall every 3 days.
a. What is the probability of five snowfalls in 2 weeks?
b. Find the probability of a snowfall today.

7.114 The number of students who seek assistance with their statistics assignments is Poisson distributed with a mean of two per day.
a. What is the probability that no students seek assistance tomorrow?
b. Find the probability that 10 students seek assistance in a week.

7.115 Hits on a personal website occur quite infrequently. They occur randomly and independently with an average of five per week.
a. Find the probability that the site gets 10 or more hits in a week.
b. Determine the probability that the site gets 20 or more hits in 2 weeks.

7.116 In older cities across North America, infrastructure is deteriorating, including water lines that supply homes and businesses. A report to the Toronto city council stated that there are on average 30 water line breaks per 100 kilometers per year in the city of Toronto. Outside of Toronto, the average number of breaks is 15 per 100 kilometers per year.
a. Find the probability that in a stretch of 100 kilometers in Toronto there are 35 or more breaks next year.
b. Find the probability that there are 12 or fewer breaks in a stretch of 100 kilometers outside of Toronto next year.

7.117 The number of bank robberies that occur in a large North American city is Poisson distributed with a mean of 1.8 per day. Find the probabilities of the following events.
a. Three or more bank robberies in a day
b. Between 10 and 15 (inclusive) robberies during a 5-day period

7.118 Flaws in a carpet tend to occur randomly and independently at a rate of one every 200 square feet. What is the probability that a carpet that is 8 feet by 10 feet contains no flaws?

7.119 Complaints about an Internet brokerage firm occur at a rate of five per day. The number of complaints appears to be Poisson distributed.
a. Find the probability that the firm receives 10 or more complaints in a day.
b. Find the probability that the firm receives 25 or more complaints in a 5-day period.

APPLICATIONS in OPERATIONS MANAGEMENT

© Lester Lefkowitz/Getty Images

Waiting Lines

Everyone is familiar with waiting lines. We wait in line at banks, groceries, and fast-food restaurants. There are also waiting lines in firms where trucks wait to load and unload and on assembly lines where stations wait for new parts. Management scientists have developed mathematical models that allow managers to determine the operating characteristics of waiting lines. Some of the operating characteristics are

The probability that there are no units in the system
The average number of units in the waiting line
The average time a unit spends in the waiting line
The probability that an arriving unit must wait for service

The Poisson probability distribution is used extensively in waiting-line (also called *queuing*) models. Many models assume that the arrival of units for service is Poisson distributed with a specific value of μ. In the next chapter, we will discuss the operating characteristics of waiting lines.

Exercises 7.120–7.122 require the calculation of the probability of a number of arrivals.

7.120 The number of trucks crossing at the Ambassador Bridge connecting Detroit, Michigan, and Windsor, Ontario, is Poisson distributed with a mean of 1.5 per minute.

a. What is the probability that in any 1-minute time span two or more trucks will cross the bridge?

b. What is the probability that fewer than four trucks will cross the bridge over the next 4 minutes?

7.121 Cars arriving for gasoline at a particular gas station follow a Poisson distribution with a mean of 5 per hour.

a. Determine the probability that over the next hour only one car will arrive.

b. Compute the probability that in the next 3 hours more than 20 cars will arrive.

7.122 The number of users of an automatic banking machine is Poisson distributed. The mean number of users per 5-minute interval is 1.5. Find the probability of the following events.

a. No users in the next 5 minutes

b. Five or fewer users in the next 15 minutes

c. Three or more users in the next 10 minutes

CHAPTER SUMMARY

There are two types of random variables. A **discrete random variable** is one whose values are countable. A **continuous random variable** can assume an uncountable number of values. In this chapter, we discussed discrete random variables and their **probability distributions**. We defined the **expected value**, **variance**, and **standard deviation** of a population represented by a discrete probability distribution. Also introduced in this chapter were **bivariate discrete distributions** on which an important application in finance was based. Finally, the two most important discrete distributions—the **binomial** and the **Poisson**—were presented.

IMPORTANT TERMS

Random variable 218
Discrete random variable 219
Continuous random variable 219
Probability distribution 219
Expected value 223
Bivariate distribution 229
PERT (Project Evaluation and Review Technique) 235
CPM (Critical Path Method) 235
Path 235
Critical path 236
Diversification 237
Binomial experiment 244
Bernoulli process 244
Binomial random variable 244
Binomial probability distribution 245
Cumulative probability 247
Poisson distribution 251
Poisson random variable 251
Poisson experiment 251

SYMBOLS

Symbol	Pronounced	Represents
$\sum_{\text{all } x} x$	Sum of x for all values of x	Summation
C_x^n	n choose x	Number of combinations
$n!$	n factorial	$n(n-1)(n-2)\cdots(3)(2)(1)$
e		$2.71828\ldots$

FORMULAS

Expected value (mean)

$$E(X) = \mu = \sum_{\text{all } x} xP(x)$$

Variance

$$V(x) = \sigma^2 = \sum_{\text{all } x} (x-\mu)^2 P(x)$$

Standard deviation

$$\sigma = \sqrt{\sigma^2}$$

Covariance

$$\text{COV}(X, Y) = \sigma_{xy} = \sum (x-\mu_x)(y-\mu_y)P(x, y)$$

Coefficient of Correlation

$$\rho = \frac{\text{COV}(X, Y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Laws of expected value

1. $E(c) = c$
2. $E(X + c) = E(X) + c$
3. $E(cX) = cE(X)$

Laws of variance

1. $V(c) = 0$
2. $V(X + c) = V(X)$
3. $V(cX) = c^2 V(X)$

Laws of expected value and variance of the sum of two variables

1. $E(X + Y) = E(X) + E(Y)$
2. $V(X + Y) = V(X) + V(Y) + 2\text{COV}(X, Y)$

Laws of expected value and variance for the sum of k variables, where $k \geq 2$

1. $E(X_1 + X_2 + \cdots + X_k) = E(X_1) + E(X_2) + \cdots + E(X_k)$
2. $V(X_1 + X_2 + \cdots + X_k) = V(X_1) + V(X_2) + \cdots + V(X_k)$

   if the variables are independent

Mean and variance of a portfolio of two stocks

$$E(Rp) = w_1 E(R_1) + w_2 E(R_2)$$

$$\begin{aligned} V(R_p) &= w_1^2 V(R_1) + w_2^2 V(R_2) \\ &\quad + 2w_1 w_2 \text{COV}(R_1, R_2) \\ &= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2w_1 w_2 \rho \sigma_1 \sigma_2 \end{aligned}$$

Mean and variance of a portfolio of k stocks

$$E(R_p) = \sum_{i=1}^{k} w_i E(R_i)$$

$$V(R_p) = \sum_{i=1}^{k} w_i^2 \sigma_i^2 + 2\sum_{i=1}^{k} \sum_{j=i+1}^{k} w_i w_j \text{COV}(R_i, R_j)$$

Binomial probability

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

$$\sigma = \sqrt{np(1-p)}$$

Poisson probability

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

COMPUTER INSTRUCTIONS

Probability Distribution	Excel	Minitab
Binomial	248	249
Poisson	255	255

CHAPTER EXERCISES

7.123 In 2000, Northwest Airlines boasted that 77.4% of its flights were on time. If we select five Northwest flights at random, what is the probability that all five are on time? (*Source*: Department of Transportation.)

7.124 The final exam in a one-term statistics course is taken in the December exam period. Students who are sick or have other legitimate reasons for missing the exam are allowed to write a deferred exam scheduled for the first week in January. A statistics professor has observed that only 2% of all students legitimately miss the December final exam. Suppose that the professor has 40 students registered this term.

a. How many students can the professor expect to miss the December exam?
b. What is the probability that the professor will not have to create a deferred exam?

7.125 The number of magazine subscriptions per household is represented by the following probability distribution.

Magazine subscriptions per household	0	1	2	3	4
Probability	.48	.35	.08	.05	.04

a. Calculate the mean number of magazine subscriptions per household.
b. Find the standard deviation.

7.126 The number of arrivals at a car wash is Poisson distributed with a mean of eight per hour.

a. What is the probability that 10 cars will arrive in the next hour?
b. What is the probability that more than 5 cars will arrive in the next hour?
c. What is the probability that fewer than 12 cars will arrive in the next hour?

7.127 The percentage of customers who enter a restaurant and ask to be seated in a smoking section is 15%. Suppose that 100 people enter the restaurant.

a. What is the expected number of people who request a smoking table?
b. What is the standard deviation of the number of requests for a smoking table?
c. What is the probability that 20 or more people request a smoking table?

7.128 Lotteries are an important income source for various governments around the world. However, the availability of lotteries and other forms of gambling have created a social problem: gambling addicts. A critic of government-controlled gambling contends that 30% of people who regularly buy lottery tickets are gambling addicts. If we randomly select 10 people among those who report that they regularly buy lottery tickets, what is the probability that more than 5 of them are addicts?

7.129 The distribution of the number of home runs in soft-ball games is shown here.

Number of home runs	0	1	2	3	4	5
Probability	.05	.16	.41	.27	.07	.04

a. Calculate the mean number of home runs.
b. Find the standard deviation.

7.130 An auditor is preparing for a physical count of inventory as a means of verifying its value. Items counted are reconciled with a list prepared by the storeroom supervisor. In one particular firm, 20% of the items counted cannot be reconciled without reviewing invoices. The auditor selects 10 items. Find the probability that 6 or more items cannot be reconciled.

7.131 Shutouts in the National Hockey League occur randomly and independently at a rate of 1 every 20 games. Calculate the probability of the following events.

a. 2 shutouts in the next 10 games
b. 25 shutouts in 400 games
c. a shutout in to.night's game

7.132 Most Miami Beach restaurants offer "early-bird" specials. These are lower-priced meals that are available only from 4 to 6 P.M. However, not all customers who arrive between 4 and 6 P.M. order the special. In fact, only 70% do.

a. Find the probability that of 80 customers between 4 and 6 P.M., more than 65 order the special.
b. What is the expected number of customers who order the special?
c. What is the standard deviation?

7.133 According to climatologists, the long-term average for Atlantic storms is 9.6 per season (June 1 to November 30), with 6 becoming hurricanes and 2.3 becoming intense hurricanes. Find the probability of the following events.

a. Ten or more Atlantic storms
b. Five or fewer hurricanes
c. Three or more intense hurricanes

7.134 Researchers at the University of Pennsylvania School of Medicine theorized that children under 2 years old who sleep in rooms with the light on have a 40% probability of becoming myopic by age 16. Suppose that researchers found 25 children who slept with the light on before they were 2.

a. What is the probability that 10 of them will become myopic before age 16?
b. What is the probability that fewer than 5 of them will become myopic before age 16?
c. What is the probability that more than 15 of them will become myopic before age 16?

7.135 A pharmaceutical researcher working on a cure for baldness noticed that middle-aged men who are balding at the crown of their head have a 45% probability of suffering a heart attack over the next decade. In a sample of 100 middle-age balding men, what are the following probabilities?

a. More than 50 will suffer a heart attack in the next decade.
b. Fewer than 44 will suffer a heart attack in the next decade.
c. Exactly 45 will suffer a heart attack in the next decade.

7.136 Advertising researchers have developed a theory that states that commercials that appear in violent television shows are less likely to be remembered and will thus be less effective. After examining samples of viewers who watch violent and nonviolent programs and asking them a series of five questions about the commercials, the researchers produced the following probability distributions of the number of correct answers.

Viewers of violent shows

x	0	1	2	3	4	5
$P(x)$	.36	.22	.20	.09	.08	.05

Viewers of nonviolent shows

x	0	1	2	3	4	5
$P(x)$	.15	.18	.23	.26	.10	.08

a. Calculate the mean and standard deviation of the number of correct answers among viewers of violent television programs.
b. Calculate the mean and standard deviation of the number of correct answers among viewers of nonviolent television programs.

7.137 According to the U.S. census, one-third of all businesses are owned by women. If we select 25 businesses at random, what is the probability that 10 or more of them are owned by women?

7.138 It is recommended that women age 40 and older have a mammogram annually. A recent report indicated that if a woman has annual mammograms over a 10-year period, there is a 60% probability that there will be at least one false-positive result. (A false-positive mammogram test result is one that indicates the presence of cancer when, in fact, there is no cancer.) If the annual test results are independent, what is the probability that in any one year a mammogram will produce a false-positive result? (*Hint:* Find the value of p such that the probability that a binomial random variable with $n = 10$ is greater than or equal to 1 is .60.)

7.139 In a recent election, the mayor received 60% of the vote. Last week, a survey was undertaken that asked 100 people whether they would vote for the mayor. Assuming that her popularity has not changed, what is the probability that more than 50 people in the sample would vote for the mayor?

7.140 When Earth traveled through the storm of meteorites trailing the comet Tempel-Tuttle on November 17, 1998, the storm was 1,000 times as intense as the average meteor storm. Before the comet arrived, telecommunication companies worried about the potential damage that might be inflicted on the approximately 650 satellites in orbit. It was estimated that each satellite had a 1% chance of being hit, causing damage to the satellite's electronic system. One company had five satellites in orbit at the time. Determine the probability distribution of the number of the company's satellites that would be damaged.

CASE 7.1 To Bunt or Not to Bunt, That Is the Question—Part 2

© AP Photo/Gene J. Puskar

In Case 6.2, we presented the probabilities of scoring at least one run and asked you to determine whether the manager should signal for the batter to sacrifice bunt. The decision was made on the basis of comparing the probability of scoring at least one run when the manager signaled for the bunt and when he signaled the batter to swing away. Another factor that should be incorporated into the decision is the *number* of runs the manager expects his team to score. In the same article referred to in Case 6.2, the author also computed the expected number of runs scored for each situation. Table 1 lists the expected number of runs in situations that are defined by the number of outs and the bases occupied.

TABLE 1 Expected Number of Runs Scored

Bases Occupied	0 Out	1 Out	2 Outs
Bases empty	.49	.27	.10
First base	.85	.52	.23
Second base	1.06	.69	.34
Third base	1.21	.82	.38
First base and second base	1.46	1.00	.48
First base and third base	1.65	1.10	.51
Second base and third base	1.94	1.50	.62
Bases loaded	2.31	1.62	.82

Assume that the manager wishes to score as many runs as possible. Using the same probabilities of the four outcomes of a bunt listed in Case 6.2, determine whether the manager should signal the batter to sacrifice bunt.

This page intentionally left blank