6

Discrete Probability Distributions

GOALS

When you have completed this chapter, you will be able to:

1 Define the terms *probability distribution* and *random variable.*

2 Distinguish between *discrete* and *continuous probability distributions.*

3 Calculate the mean, variance, and standard deviation of a discrete probability distribution.

4 Describe the characteristics of and compute probabilities using the *binomial probability distribution.*

5 Describe the characteristics of and compute probabilities using the *hypergeometric probability distribution.*

6 Describe the characteristics of and compute probabilities using the *Poisson probability distribution.*

Croissant Bakery, Inc. offers special decorated cakes for birthdays, weddings and other occasions. They also have regular cakes available in their bakery. Based on the data provided in the table, compute the mean, variance, and standard deviation of the number of cakes sold per day. (See Exercise 44, Goal 3.)

Introduction

Chapters 2 through 4 are devoted to descriptive statistics. We describe raw data by organizing it into a frequency distribution and portraying the distribution in tables, graphs, and charts. Also, we compute a measure of location—such as the arithmetic mean, median, or mode—to locate a typical value near the center of the distribution. The range and the standard deviation are used to describe the spread in the data. These chapters focus on describing *something that has already happened.*

Starting with Chapter 5, the emphasis changes—we begin examining *something that would probably happen.* We note that this facet of statistics is called *statistical inference.* The objective is to make inferences (statements) about a population based on a number of observations, called a sample, selected from the population. In Chapter 5, we state that a probability is a value between 0 and 1 inclusive, and we examine how probabilities can be combined using rules of addition and multiplication.

This chapter will begin the study of **probability distributions.** A probability distribution gives the entire range of values that can occur based on an experiment. A probability distribution is similar to a relative frequency distribution. However, instead of describing the past, it describes how likely some future event is. For example, a drug manufacturer may claim a treatment will cause weight loss for 80 percent of the population. A consumer protection agency may test the treatment on a sample of six people. If the manufacturer's claim is true, it is *almost impossible* to have an outcome where no one in the sample loses weight and it is *most likely* that 5 out of the 6 do lose weight.

In this chapter we discuss the mean, variance, and standard deviation of a probability distribution. We also discuss three frequently occurring probability distributions: the binomial, hypergeometric, and Poisson.

What Is a Probability Distribution?

A probability distribution shows the possible outcomes of an experiment and the probability of each of these outcomes.

PROBABILITY DISTRIBUTION A listing of all the outcomes of an experiment and the probability associated with each outcome.

How can we generate a probability distribution?

Example

Suppose we are interested in the number of heads showing face up on three tosses of a coin. This is the experiment. The possible results are: zero heads, one head, two heads, and three heads. What is the probability distribution for the number of heads?

Solution

There are eight possible outcomes. A tail might appear face up on the first toss, another tail on the second toss, and another tail on the third toss of the coin. Or we might get a tail, tail, and head, in that order. We use the multiplication formula for counting outcomes (5–8). There are (2)(2)(2) or 8 possible results. These results are listed below.

Possible Result	Coin Toss First	Second	Third	Number of Heads
1	T	T	T	0
2	T	T	H	1
3	T	H	T	1
4	T	H	H	2
5	H	T	T	1
6	H	T	H	2
7	H	H	T	2
8	H	H	H	3

Note that the outcome "zero heads" occurred only once, "one head" occurred three times, "two heads" occurred three times, and the outcome "three heads" occurred only once. That is, "zero heads" happened one out of eight times. Thus, the probability of zero heads is one-eighth, the probability of one head is three-eighths, and so on. The probability distribution is shown in Table 6–1. Because one of these outcomes must happen, the total of the probabilities of all possible events is 1.000. This is always true. The same information is shown in Chart 6–1.

TABLE 6–1 Probability Distribution for the Events of Zero, One, Two, and Three Heads Showing Face Up on Three Tosses of a Coin

Number of Heads, x	Probability of Outcome, $P(x)$
0	$\frac{1}{8} = .125$
1	$\frac{3}{8} = .375$
2	$\frac{3}{8} = .375$
3	$\frac{1}{8} = .125$
Total	$\frac{8}{8} = 1.000$

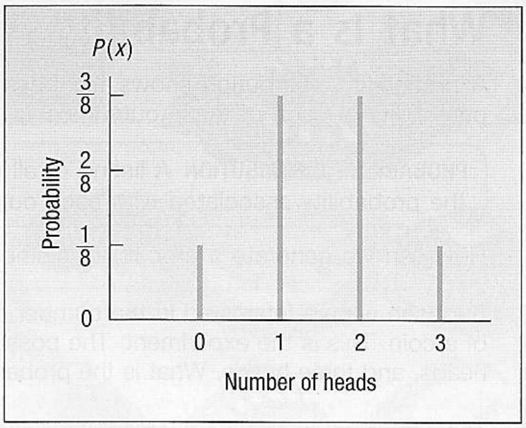


CHART 6–1 Graphical Presentation of the Number of Heads Resulting from Three Tosses of a Coin and the Corresponding Probability

Before continuing, we should note the important characteristics of a probability distribution.

CHARACTERISTICS OF A PROBABILITY DISTRIBUTION

1. The probability of a particular outcome is between 0 and 1 inclusive.
2. The outcomes are mutually exclusive events.
3. The list is exhaustive. So the sum of the probabilities of the various events is equal to 1.

Refer to the coin-tossing example in Table 6–1. We write the probability of x as $P(x)$. So the probability of zero heads is $P(0 \text{ heads}) = .125$ and the probability of one head is $P(1 \text{ head}) = .375$ and so forth. The sum of these mutually exclusive probabilities is 1; that is, from Table 6–1, $0.125 + 0.375 + 0.375 + 0.125 = 1.00$.

Self-Review 6–1

The possible outcomes of an experiment involving the roll of a six-sided die are a one-spot, a two-spot, a three-spot, a four-spot, a five-spot, and a six-spot.

(a) Develop a probability distribution for the number of possible spots.
(b) Portray the probability distribution graphically.
(c) What is the sum of the probabilities?

Random Variables

In any experiment of chance, the outcomes occur randomly. So it is often called a *random variable.* For example, rolling a single die is an experiment: any one of six possible outcomes can occur. Some experiments result in outcomes that are quantitative (such as dollars, weight, or number of children), and others result in qualitative outcomes (such as color or religious preference). Each value of the random variable is associated with a probability to indicate the chance of a particular outcome. A few examples will further illustrate what is meant by a **random variable.**

- If we count the number of employees absent from the day shift on Monday, the number might be 0, 1, 2, 3, . . . The number absent is the random variable.
- If we weigh four steel ingots, the weights might be 2,492 pounds, 2,497 pounds, 2,506 pounds, and so on. The weight is the random variable.
- If we toss two coins and count the number of heads, there could be zero, one, or two heads. Because the number of heads resulting from this experiment is due to chance, the number of heads appearing is the random variable.
- Other random variables might be the number of defective lightbulbs produced in an hour at the Cleveland Company, Inc., the grade level (9, 10, 11, or 12) of the members of the St. James girls' basketball team, the number of runners in the Boston Marathon for the 2006 race, and the daily number of drivers charged with driving under the influence of alcohol in Texas.

RANDOM VARIABLE A quantity resulting from an experiment that, by chance, can assume different values.

The following diagram illustrates the terms *experiment, outcome, event,* and *random variable.* First, for the experiment where a coin is tossed three times, there are eight possible outcomes. In this experiment, we are interested in the event that one head occurs in the three tosses. The random variable is the number of heads. In terms of probability, we want to know the probability of the event that the random variable equals 1. The result is $P(1 \text{ head in 3 tosses}) = 0.375$.

Possible *outcomes* for three coin tosses

The *event* {one head} occurs and the *random variable* $x = 1$.

A random variable may be either *discrete* or *continuous.*

Discrete Random Variable

A discrete random variable can assume only a certain number of separated values. If there are 100 employees, then the count of the number absent on Monday can only be 0, 1, 2, 3, . . . , 100. A discrete random variable is usually the result of counting something. By way of definition:

DISCRETE RANDOM VARIABLE A random variable that can assume only certain clearly separated values.

A discrete random variable can, in some cases, assume fractional or decimal values. These values must be separated, that is, have distance between them. As an example, the scores awarded by judges for technical competence and artistic form in figure skating are decimal values, such as 7.2, 8.9, and 9.7. Such values are discrete because there is distance between scores of, say, 8.3 and 8.4. A score cannot be 8.34 or 8.347, for example.

Continuous Random Variable

On the other hand, if the random variable is continuous, then the distribution is a continuous probability distribution. If we measure something such as the width of a room, the height of a person, or the pressure in an automobile tire, the variable is a *continuous random variable.* It can assume one of an infinitely large number of values, within certain limitations. As examples:

- The times of commercial flights between Atlanta and Los Angeles are 4.67 hours, 5.13 hours, and so on. The random variable is the number of hours.
- Tire pressure, measured in pounds per square inch (psi), for a new Chevy Trailblazer might be 32.78 psi, 31.62 psi, 33.07 psi, and so on. In other words, any values between 28 and 35 could reasonably occur. The random variable is the tire pressure.

Logically, if we organize a set of possible values from a random variable into a probability distribution, the result is a **probability distribution.** So what is the difference between a probability distribution and a random variable? A random variable reports the particular outcome of an experiment. A probability distribution reports all the possible outcomes as well as the corresponding probability.

The tools used, as well as the probability interpretations, are different for discrete and continuous probability distributions. This chapter is limited to the discussion and interpretation of discrete distributions. In the next chapter we discuss continuous distributions. How do you tell the difference between the two types of distributions? Usually a discrete distribution is the result of counting something, such as:

- The number of heads appearing when a coin is tossed 3 times.
- The number of students earning an A in this class.
- The number of production employees absent from the second shift today.
- The number of 30-second commercials on NBC from 8 to 11 P.M. tonight.

Continuous distributions are usually the result of some type of measurement, such as:

- The length of each song on the latest Tim McGraw album.
- The weight of each student in this class.

- The temperature outside as you are reading this book.
- The amount of money earned by each of the more than 750 players currently on Major League Baseball team rosters.

The Mean, Variance, and Standard Deviation of a Probability Distribution

In Chapter 3 we discussed measures of location and variation for a frequency distribution. The mean reports the central location of the data, and the variance describes the spread in the data. In a similar fashion, a probability distribution is summarized by its mean and variance. We identify the mean of a probability distribution by the lowercase Greek letter mu (μ) and the standard deviation by the lowercase Greek letter sigma (σ).

Mean

The mean is a typical value used to represent the central location of a probability distribution. It also is the long-run average value of the random variable. The mean of a probability distribution is also referred to as its **expected value.** It is a weighted average where the possible values of a random variable are weighted by their corresponding probabilities of occurrence.

The mean of a discrete probability distribution is computed by the formula:

MEAN OF A PROBABILITY DISTRIBUTION $$\mu = \Sigma[xP(x)]$$ **[6–1]**

where $P(x)$ is the probability of a particular value x. In other words, multiply each x value by its probability of occurrence, and then add these products.

Variance and Standard Deviation

As noted, the mean is a typical value used to summarize a discrete probability distribution. However, it does not describe the amount of spread (variation) in a distribution. The variance does this. The formula for the variance of a probability distribution is:

VARIANCE OF A PROBABILITY DISTRIBUTION $$\sigma^2 = \Sigma[(x - \mu)^2P(x)]$$ **[6–2]**

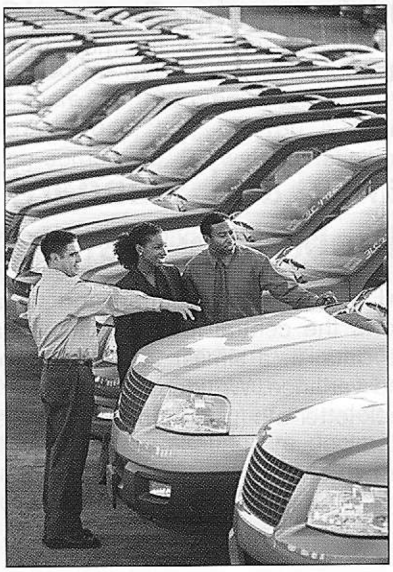
The computational steps are

1. Subtract the mean from each value, and square this difference.
2. Multiply each squared difference by its probability.
3. Sum the resulting products to arrive at the variance.

The standard deviation, σ, is found by taking the positive square root of σ^2; that is, $\sigma = \sqrt{\sigma^2}$.

An example will help explain the details of the calculation and interpretation of the mean and standard deviation of a probability distribution.

Example



John Ragsdale sells new cars for Pelican Ford. John usually sells the largest number of cars on Saturday. He has developed the following probability distribution for the number of cars he expects to sell on a particular Saturday.

Number of Cars Sold, *x*	Probability, *P*(*x*)
0	.10
1	.20
2	.30
3	.30
4	.10
Total	1.00

1. What type of distribution is this?
2. On a typical Saturday, how many cars does John expect to sell?
3. What is the variance of the distribution?

Solution

1. This is a discrete probability distribution for the random variable called "number of cars sold." Note that John expects to sell only within a certain range of cars; he does not expect to sell 5 cars or 50 cars. Further, he cannot sell half a car. He can sell only 0, 1, 2, 3, or 4 cars. Also, the outcomes are mutually exclusive—he cannot sell a total of both 3 and 4 cars on the same Saturday.
2. The mean number of cars sold is computed by weighting the number of cars sold by the probability of selling that number and adding or summing the products, using formula (6–1):

$$\mu = \Sigma[xP(x)]$$
$$= 0(.10) + 1(.20) + 2(.30) + 3(.30) + 4(.10)$$
$$= 2.1$$

These calculations are summarized in the following table.

Number of Cars Sold, *x*	Probability, *P*(*x*)	*x* · *P*(*x*)
0	.10	0.00
1	.20	0.20
2	.30	0.60
3	.30	0.90
4	.10	0.40
Total	1.00	μ = 2.10

How do we interpret a mean of 2.1? This value indicates that, over a large number of Saturdays, John Ragsdale expects to sell a mean of 2.1 cars a day. Of course, it is not possible for him to sell *exactly* 2.1 cars on any particular Saturday. However, the expected value can be used to predict the arithmetic mean number of cars sold on Saturdays in the long run. For example, if John works 50 Saturdays during a year, he can expect to sell (50)(2.1) or 105 cars just on Saturdays. Thus, the mean is sometimes called the expected value.

3. Again, a table is useful for systemizing the computations for the variance, which is 1.290.

Number of Cars Sold, x	Probability, $P(x)$	$(x - \mu)$	$(x - \mu)^2$	$(x - \mu)^2 P(x)$
0	.10	0 − 2.1	4.41	0.441
1	.20	1 − 2.1	1.21	0.242
2	.30	2 − 2.1	0.01	0.003
3	.30	3 − 2.1	0.81	0.243
4	.10	4 − 2.1	3.61	0.361
				$\sigma^2 = 1.290$

Recall that the standard deviation, σ, is the positive square root of the variance. In this example, $\sqrt{\sigma^2} = \sqrt{1.290} = 1.136$ cars. How do we interpret a standard deviation of 1.136 cars? If salesperson Rita Kirsch also sold a mean of 2.1 cars on Saturdays, and the standard deviation in her sales was 1.91 cars, we would conclude that there is more variability in the Saturday sales of Ms. Kirsch than in those of Mr. Ragsdale (because $1.91 > 1.136$).

Self-Review 6–2

The Pizza Palace offers three sizes of cola—small, medium, and large—to go with its pizza.

The colas are sold for $0.80, $0.90, and $1.20, respectively. Thirty percent of the orders are for small, 50 percent are for medium, and 20 percent are for the large sizes. Organize the size of the colas and the probability of a sale into a probability distribution.

(a) Is this a discrete probability distribution? Indicate why or why not.
(b) Compute the mean amount charged for a cola.
(c) What is the variance in the amount charged for a cola? The standard deviation?

Exercises

1. Compute the mean and variance of the following discrete probability distribution.

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

2. Compute the mean and variance of the following discrete probability distribution.

x	$P(x)$
2	.5
8	.3
10	.2

3. Three tables listed at the top of page 188 show "random variables" and their "probabilities." However, only one of these is actually a probability distribution.

a. Which is it?

x	$P(x)$
5	.3
10	.3
15	.2
20	.4

x	$P(x)$
5	.1
10	.3
15	.2
20	.4

x	$P(x)$
5	.5
10	.3
15	−.2
20	.4

b. Using the correct probability distribution, find the probability that x is:
(1) Exactly 15. **(2)** No more than 10. **(3)** More than 5.

c. Compute the mean, variance, and standard deviation of this distribution.

4. Which of these variables are discrete and which are continuous random variables?
a. The number of new accounts established by a salesperson in a year.
b. The time between customer arrivals to a bank ATM.
c. The number of customers in Big Nick's barber shop.
d. The amount of fuel in your car's gas tank.
e. The number of minorities on a jury.
f. The outside temperature today.

5. The information below is the number of daily emergency service calls made by the volunteer ambulance service of Walterboro, South Carolina, for the last 50 days. To explain, there were 22 days on which there were 2 emergency calls, and 9 days on which there were 3 emergency calls.

Number of Calls	Frequency
0	8
1	10
2	22
3	9
4	1
Total	50

a. Convert this information on the number of calls to a probability distribution.
b. Is this an example of a discrete or continuous probability distribution?
c. What is the mean number of emergency calls per day?
d. What is the standard deviation of the number of calls made daily?

6. The director of admissions at Kinzua University in Nova Scotia estimated the distribution of student admissions for the fall semester on the basis of past experience. What is the expected number of admissions for the fall semester? Compute the variance and the standard deviation of the number of admissions.

Admissions	Probability
1,000	.6
1,200	.3
1,500	.1

7. Belk Department Store is having a special sale this weekend. Customers charging purchases of more than $50 to their Belk credit card will be given a special Belk Lottery card. The customer will scratch off the card, which will indicate the amount to be taken off the total amount of the purchase. Listed below are the amount of the prize and the percent of the time that amount will be deducted from the total amount of the purchase.

Prize Amount	Probability
$ 10	.50
25	.40
50	.08
100	.02

a. What is the mean amount deducted from the total purchase amount?
b. What is the standard deviation of the amount deducted from the total purchase?

8. The Downtown Parking Authority of Tampa, Florida, reported the following information for a sample of 250 customers on the number of hours cars are parked and the amount they are charged.

Number of Hours	Frequency	Amount Charged
1	20	$ 3.00
2	38	6.00
3	53	9.00
4	45	12.00
5	40	14.00
6	13	16.00
7	5	18.00
8	36	20.00
	250	

a. Convert the information on the number of hours parked to a probability distribution. Is this a discrete or a continuous probability distribution?
b. Find the mean and the standard deviation of the number of hours parked. How would you answer the question: How long is a typical customer parked?
c. Find the mean and the standard deviation of the amount charged.

Binomial Probability Distribution

The **binomial probability distribution** is a widely occurring discrete probability distribution. One characteristic of a binomial distribution is that there are only two possible outcomes on a particular trial of an experiment. For example, the statement in a true/false question is either true or false. The outcomes are mutually exclusive, meaning that the answer to a true/false question cannot be both true and false at the same time. As other examples, a product is classified as either acceptable or not acceptable by the quality control department, a worker is classified as employed or unemployed, and a sales call results in the customer either purchasing the product or not purchasing the product. Frequently, we classify the two possible outcomes as "success" and "failure." However, this classification does *not* imply that one outcome is good and the other is bad.

Another characteristic of the binomial distribution is that the random variable is the result of counts. That is, we count the number of successes in the total number of trials. We flip a fair coin five times and count the number of times a head appears, we select 10 workers and count the number who are over 50 years of age, or we select 20 boxes of Kellogg's Raisin Bran and count the number that weigh more than the amount indicated on the package.

A third characteristic of a binomial distribution is that the probability of a success remains the same from one trial to another. Two examples are:

- The probability you will guess the first question of a true/false test correctly (a success) is one-half. This is the first "trial." The probability that you will guess correctly on the second question (the second trial) is also one-half, the probability of success on the third trial is one-half, and so on.

- If past experience revealed the swing bridge over the Intracoastal Waterway in Socastee was raised one out of every 20 times you approach it, then the probability is one-twentieth that it will be raised (a "success") the next time you approach it, one-twentieth the following time, and so on.

The final characteristic of a binomial probability distribution is that each trial is *independent* of any other trial. Independent means that there is no pattern to the trials. The outcome of a particular trial does not affect the outcome of any other trial.

Binomial characteristics

BINOMIAL PROBABILITY EXPERIMENT

1. An outcome on each trial of an experiment is classified into one of two mutually exclusive categories—a success or a failure.
2. The random variable counts the number of successes in a fixed number of trials.
3. The probability of success and failure stay the same for each trial.
4. The trials are independent, meaning that the outcome of one trial does not affect the outcome of any other trial.

How Is a Binomial Probability Computed?

To construct a particular binomial probability, we use (1) the number of trials and (2) the probability of success on each trial. For example, if an examination at the conclusion of a management seminar consists of 20 multiple-choice questions, the number of trials is 20. If each question has five choices and only one choice is correct, the probability of success on each trial is .20. Thus, the probability is .20 that a person with no knowledge of the subject matter will guess the answer to a question correctly. So the conditions of the binomial distribution just noted are met.

A binomial probability is computed by the formula:

BINOMIAL PROBABILITY FORMULA $$P(x) = {}_nC_x\, \pi^x(1 - \pi)^{n-x} \qquad [6\text{–}3]$$

where:

C denotes a combination.
n is the number of trials.
x is the random variable defined as the number of successes.
π is the probability of a success on each trial.

We use the Greek letter π (pi) to denote a binomial population parameter. Do not confuse it with the mathematical constant 3.1416.

Example

There are five flights daily from Pittsburgh via US Airways into the Bradford, Pennsylvania, Regional Airport. Suppose the probability that any flight arrives late is .20. What is the probability that none of the flights are late today? What is the probability that exactly one of the flights is late today?

Solution

We can use Formula (6–3). The probability that a particular flight is late is .20, so let π = .20. There are five flights, so n = 5, and x, the random variable, refers to the number of successes. In this case a "success" is a plane that arrives late. Because there are no late arrivals x = 0.

$$P(0) = {}_nC_x(\pi)^x(1 - \pi)^{n-x}$$
$$= {}_5C_0(.20)^0(1 - .20)^{5-0} = (1)(1)(.3277) = .3277$$

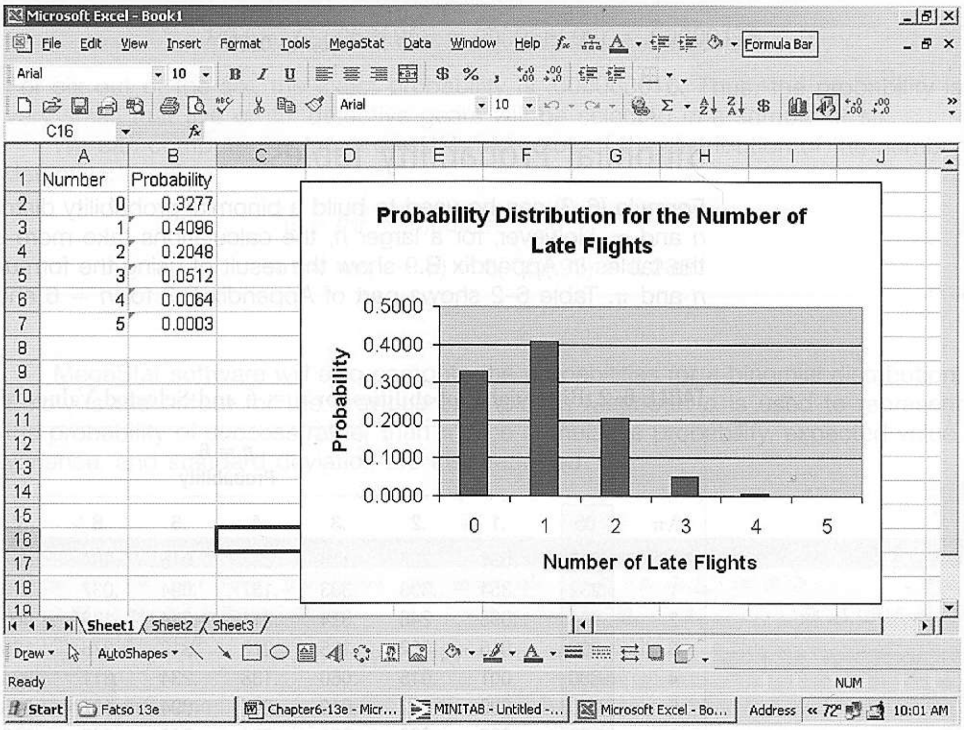
The probability that exactly one of the five flights will arrive late today is .4096, found by

$$P(1) = {}_nC_x(\pi)^x(1-\pi)^{n-x}$$

$$= {}_5C_1(.20)^1(1-.20)^{5-1} = (5)(.20)(.4096) = .4096$$

The entire binomial probability distribution with $\pi = .20$ and $n = 5$ is shown in the left portion of the following Excel spreadsheet. Also shown is a bar chart of the probability distribution. We can observe that the probability of exactly 3 late flights is .0512 and from the bar chart that the distribution of the number of late arrivals is positively skewed. The Excel instructions to compute these probabilities are the same as for the Excel output on page 194. These instructions are detailed on page 219.

Excel



The mean (μ) and the variance (σ^2) of a binomial distribution can be computed in a "shortcut" fashion by:

MEAN OF A BINOMIAL DISTRIBUTION $\mu = n\pi$ **[6–4]**

VARIANCE OF A BINOMIAL DISTRIBUTION $\sigma^2 = n\pi(1-\pi)$ **[6–5]**

For the example regarding the number of late flights, recall that $\pi = .20$ and $n = 5$. Hence:

$$\mu = n\pi = (5)(.20) = 1.0$$

$$\sigma^2 = n\pi(1-\pi) = 5(.20)(1-.20) = .80$$

The mean of 1.0 and the variance of .80 can be verified from formulas (6–1) and (6–2). The probability distribution from the Excel output on the previous page and the details of the calculations are shown below.

Number of Late Flights, x	$P(x)$	$xP(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2P(x)$
0	0.3277	0.0000	−1	1	0.3277
1	0.4096	0.4096	0	0	0
2	0.2048	0.4096	1	1	0.2048
3	0.0512	0.1536	2	4	0.2048
4	0.0064	0.0256	3	9	0.0576
5	0.0003	0.0015	4	16	0.0048
		μ = 1.0000			σ^2 = 0.7997

Binomial Probability Tables

Formula (6–3) can be used to build a binomial probability distribution for any value of n and π. However, for a larger n, the calculations take more time. For convenience, the tables in Appendix B.9 show the result of using the formula for various values of n and π. Table 6–2 shows part of Appendix B.9 for $n = 6$ and various values of π.

TABLE 6–2 Binomial Probabilities for $n = 6$ and Selected Values of π

					$n = 6$ Probability						
$x\backslash\pi$	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
0	.735	.531	.262	.118	.047	.016	.004	.001	.000	.000	.000
1	.232	.354	.393	.303	.187	.094	.037	.010	.002	.000	.000
2	.031	.098	.246	.324	.311	.234	.138	.060	.015	.001	.000
3	.002	.015	.082	.185	.276	.313	.276	.185	.082	.015	.002
4	.000	.001	.015	.060	.138	.234	.311	.324	.246	.098	.031
5	.000	.000	.002	.010	.037	.094	.187	.303	.393	.354	.232
6	.000	.000	.000	.001	.004	.016	.047	.118	.262	.531	.735

Example

Five percent of the worm gears produced by an automatic, high-speed Carter-Bell milling machine are defective. What is the probability that out of six gears selected at random none will be defective? Exactly one? Exactly two? Exactly three? Exactly four? Exactly five? Exactly six out of six?

Solution

The binomial conditions are met: (a) there are only two possible outcomes (a particular gear is either defective or acceptable), (b) there is a fixed number of trials (6), (c) there is a constant probability of success (.05), and (d) the trials are independent.

Refer to Table 6–2 above for the probability of exactly zero defective gears. Go down the left margin to an x of 0. Now move horizontally to the column headed by a π of .05 to find the probability. It is .735.

The probability of exactly one defective in a sample of six worm gears is .232. The complete binomial probability distribution for $n = 6$ and $\pi = .05$ is:

Number of Defective Gears, x	Probability of Occurrence, $P(x)$	Number of Defective Gears, x	Probability of Occurrence, $P(x)$
0	.735	4	.000
1	.232	5	.000
2	.031	6	.000
3	.002		

Of course, there is a slight chance of getting exactly five defective gears out of six random selections. It is .00000178, found by inserting the appropriate values in the binomial formula:

$$P(5) = {}_6C_5(.05)^5(.95)^1 = (6)(.05)^5(.95) = .00000178$$

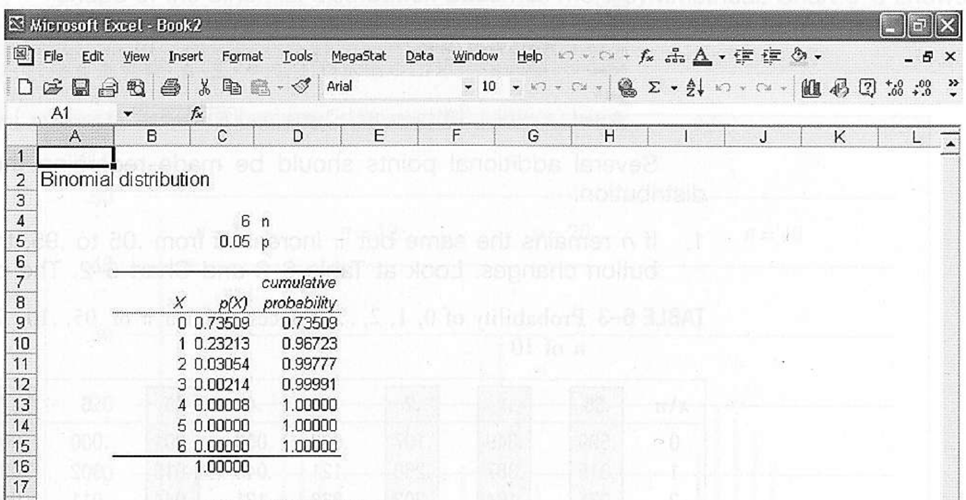
For six out of the six, the exact probability is .000000016. Thus, the probability is very small that five or six defective gears will be selected in a sample of six.

We can compute the mean or expected value of the distribution of the number defective:

$$\mu = n\pi = (6)(.05) = 0.30$$

$$\sigma^2 = n\pi(1 - \pi) = 6(.05)(.95) = 0.285$$

MegaStat software will also compute the probabilities for a binomial distribution. Below is the output for the previous example. In MegaStat p is used to represent the probability of success rather than π. The cumulative probability, expected value, variance, and standard deviation are also reported.



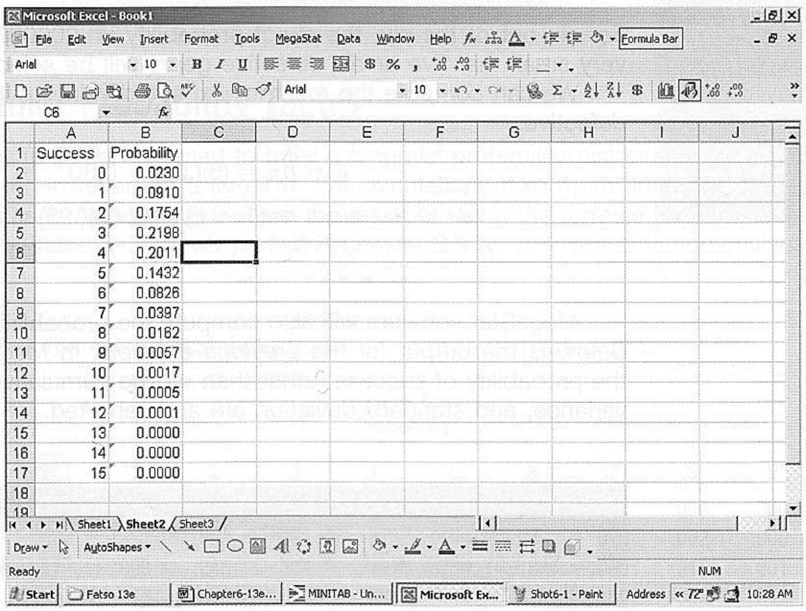
MegaStat

Self-Review 6–3 Eighty percent of the employees at the General Mills plant on Laskey Road have their bimonthly wages sent directly to their bank by electronic funds transfer. This is also called direct deposit. Suppose we select a random sample of seven recipients.

(a) Does this situation fit the assumptions of the binomial distribution?
(b) What is the probability that all seven employees use direct deposit?
(c) Use formula (6–3) to determine the exact probability that four of the seven sampled employees use direct deposit.
(d) Use Appendix B.9 to verify your answers to parts (b) and (c).

Appendix B.9 is limited. It gives probabilities for n values from 1 to 15 and π values of .05, .10, . . . , .90, and .95. A software program can generate the probabilities for a specified number of successes, given n and π. The Excel output below shows the probability when $n = 40$ and $\pi = .09$. Note that the number of successes stops at 15 because the probabilities for 16 to 40 are very close to 0.

Excel



Success	Probability
0	0.0230
1	0.0910
2	0.1754
3	0.2198
4	0.2011
5	0.1432
6	0.0826
7	0.0397
8	0.0162
9	0.0057
10	0.0017
11	0.0005
12	0.0001
13	0.0000
14	0.0000
15	0.0000

Several additional points should be made regarding the binomial probability distribution.

1. If n remains the same but π increases from .05 to .95, the shape of the distribution changes. Look at Table 6–3 and Chart 6–2. The probabilities for a π of

TABLE 6–3 Probability of 0, 1, 2, . . . Successes for a π of .05, .10, .20, .50, and .70 and an n of 10

$x \backslash \pi$	.05	.1	.2	.3	.4	.5	.6	.7	.8	.9	.95
0	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000
1	.315	.387	.268	.121	.040	.010	.002	.000	.000	.000	.000
2	.075	.194	.302	.233	.121	.044	.011	.001	.000	.000	.000
3	.010	.057	.201	.267	.215	.117	.042	.009	.001	.000	.000
4	.001	.011	.088	.200	.251	.205	.111	.037	.006	.000	.000
5	.000	.001	.026	.103	.201	.246	.201	.103	.026	.001	.000
6	.000	.000	.006	.037	.111	.205	.251	.200	.088	.011	.001
7	.000	.000	.001	.009	.042	.117	.215	.267	.201	.057	.010
8	.000	.000	.000	.001	.011	.044	.121	.233	.302	.194	.075
9	.000	.000	.000	.000	.002	.010	.040	.121	.268	.387	.315
10	.000	.000	.000	.000	.000	.001	.006	.028	.107	.349	.599

.05 are positively skewed. As π approaches .50, the distribution becomes symmetrical. As π goes beyond .50 and moves toward .95, the probability distribution becomes negatively skewed. Table 6–3 highlights probabilities for $n = 10$ and π of .05, .10, .20, .50, and .70. The graphs of these probability distributions are shown in Chart 6–2.

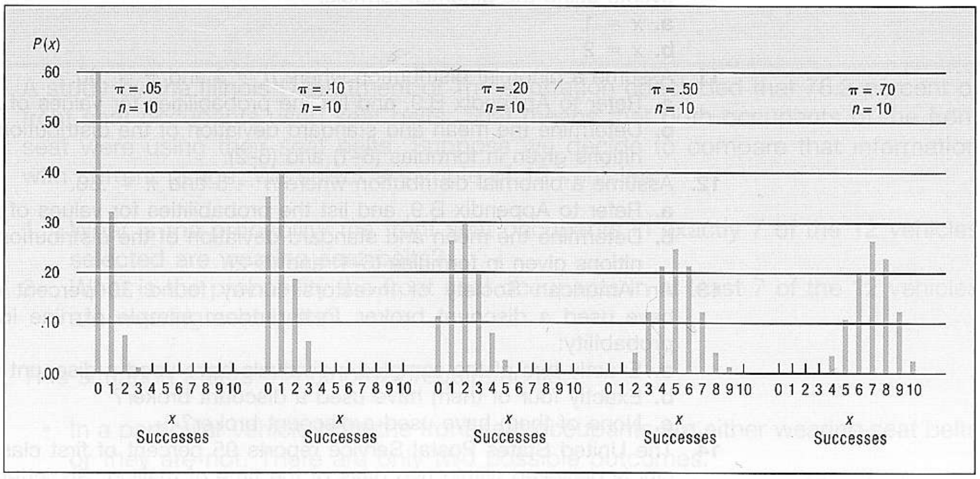


CHART 6–2 Graphing the Binomial Probability Distribution for a π of .05, .10, .20, .50, and .70 and an *n* of 10

2. If π, the probability of success, remains the same but *n* becomes larger, the shape of the binomial distribution becomes more symmetrical. Chart 6–3 shows a situation where π remains constant at .10 but *n* increases from 7 to 40.

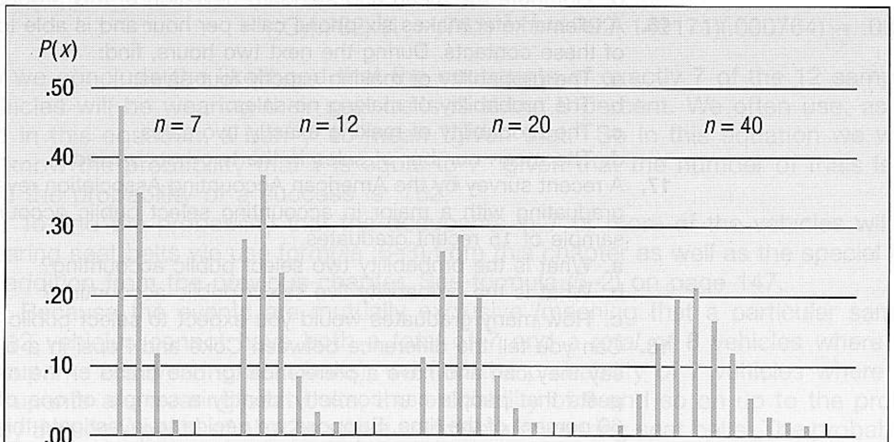


CHART 6–3 Chart Representing the Binomial Probability Distribution for a π of .10 and an *n* of 7, 12, 20, and 40

Exercises

9. In a binomial situation $n = 4$ and $\pi = .25$. Determine the probabilities of the following events using the binomial formula.
   a. $x = 2$
   b. $x = 3$
10. In a binomial situation $n = 5$ and $\pi = .40$. Determine the probabilities of the following events using the binomial formula.
    a. $x = 1$
    b. $x = 2$
11. Assume a binomial distribution where $n = 3$ and $\pi = .60$.
    a. Refer to Appendix B.9, and list the probabilities for values of x from 0 to 3.
    b. Determine the mean and standard deviation of the distribution from the general definitions given in formulas (6–1) and (6–2).
12. Assume a binomial distribution where $n = 5$ and $\pi = .30$.
    a. Refer to Appendix B.9, and list the probabilities for values of x from 0 to 5.
    b. Determine the mean and standard deviation of the distribution from the general definitions given in formulas (6–1) and (6–2).
13. An American Society of Investors survey found 30 percent of individual investors have used a discount broker. In a random sample of nine individuals, what is the probability:
    a. Exactly two of the sampled individuals have used a discount broker?
    b. Exactly four of them have used a discount broker?
    c. None of them have used a discount broker?
14. The United States Postal Service reports 95 percent of first class mail within the same city is delivered within two days of the time of mailing. Six letters are randomly sent to different locations.
    a. What is the probability that all six arrive within two days?
    b. What is the probability that exactly five arrive within two days?
    c. Find the mean number of letters that will arrive within two days.
    d. Compute the variance and standard deviation of the number that will arrive within two days.
15. Industry standards suggest that 10 percent of new vehicles require warranty service within the first year. Jones Nissan in Sumter, South Carolina, sold 12 Nissans yesterday.
    a. What is the probability that none of these vehicles requires warranty service?
    b. What is the probability exactly one of these vehicles requires warranty service?
    c. Determine the probability that exactly two of these vehicles require warranty service.
    d. Compute the mean and standard deviation of this probability distribution.
16. A telemarketer makes six phone calls per hour and is able to make a sale on 30 percent of these contacts. During the next two hours, find:
    a. The probability of making exactly four sales.
    b. The probability of making no sales.
    c. The probability of making exactly two sales.
    d. The mean number of sales in the two-hour period.
17. A recent survey by the American Accounting Association revealed 23 percent of students graduating with a major in accounting select public accounting. Suppose we select a sample of 15 recent graduates.
    a. What is the probability two select public accounting?
    b. What is the probability five select public accounting?
    c. How many graduates would you expect to select public accounting?
18. Can you tell the difference between Coke and Pepsi in a blind taste test? Most people say they can and have a preference for one brand or the other. However, research suggests that people can correctly identify a sample of one of these products only about 60 percent of the time. Suppose we decide to investigate this question and select a sample of 15 college students.
    a. How many of the 15 students would you expect to correctly identify Coke or Pepsi?
    b. What is the probability exactly 10 of the students surveyed will correctly identify Coke or Pepsi?
    c. What is the probability at least 10 of the students will correctly identify Coke or Pepsi?

Cumulative Binomial Probability Distributions

We may wish to know the probability of correctly guessing the answers to 6 *or more* true/false questions out of 10. Or we may be interested in the probability of *selecting less than two* defectives at random from production during the previous hour. In these cases we need cumulative frequency distributions similar to the ones developed in Chapter 2. See page 41. The following example will illustrate.

Example

A study by the Illinois Department of Transportation concluded that 76.2 percent of front seat occupants used seat belts. That means that both occupants of the front seat were using their seat belts. Suppose we decide to compare that information with current usage. We select a sample of 12 vehicles.

1. What is the probability the front seat occupants in exactly 7 of the 12 vehicles selected are wearing seat belts?
2. What is the probability the front seat occupants in at least 7 of the 12 vehicles are wearing seat belts?

Solution

This situation meets the binomial requirements.

- In a particular vehicle both the front seat occupants are either wearing seat belts or they are not. There are only two possible outcomes.
- There are a fixed number of trials, 12 in this case, because 12 vehicles are checked.
- The probability of a "success" (occupants wearing seat belts) is the same from one vehicle to the next: 76.2 percent.
- The trials are independent. If the fourth vehicle selected in the sample has all the occupants wearing their seat belts, this does not have any effect on the results for the fifth or tenth vehicle.

To find the likelihood the occupants of *exactly* 7 of the sampled vehicles are wearing seat belts we use formula 6-3. In this case $n = 12$ and $\pi = .762$.

$$P(x = 7 \mid n = 12 \text{ and } \pi = .762)$$
$$= {}_{12}C_7(.762)^7(1 - .762)^{12-7} = 792(.149171)(.000764) = .0902$$

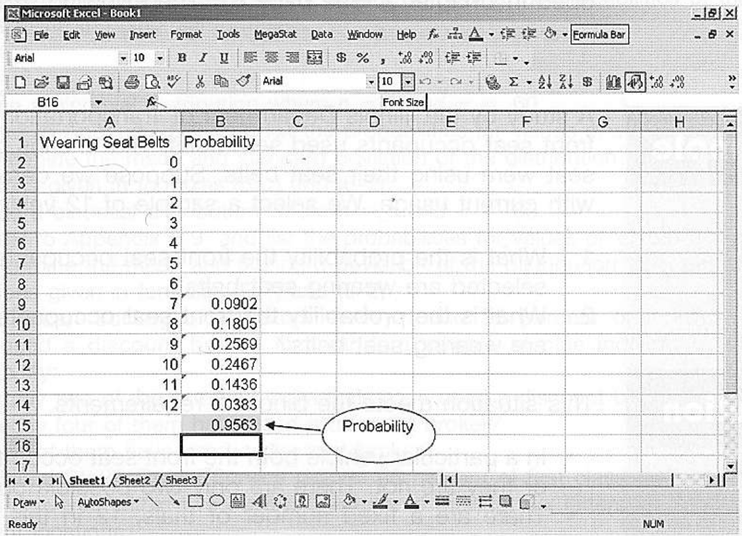
So we conclude the likelihood that the occupants of exactly 7 of the 12 sampled vehicles will be wearing their seat belts is about 9 percent. We often use, as we did in this equation, a bar "|" to mean "given that." So in this equation we want to know the probability that x is equal to 7 "given that the number of trials is 12 and the probability of a success is .762."

To find the probability that the occupants in 7 or more of the vehicles will be wearing seat belts we use formula (6–3) from this chapter as well as the special rule of addition from the previous chapter. See formula (5–2) on page 147.

Because the events are mutually exclusive (meaning that a particular sample of 12 vehicles cannot have both a *total* of 7 and a *total* of 8 vehicles where the occupants are wearing seat belts), we find the probability of 7 vehicles where the occupants are wearing seat belts, the probability of 8 and so on up to the probability that occupants of all 12 sample vehicles are wearing seat belts. The probability of each of these outcomes is then totaled.

$$P(x \geq 7 \mid n = 12 \text{ and } \pi = .762$$
$$= P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10) + P(x = 11) + P(x = 12)$$
$$= .0902 + .1805 + .2569 + .2467 + .1436 + .0383$$
$$= .9562$$

So the probability of selecting 12 cars and finding that the occupants of 7 or more vehicles were wearing seat belts is .9562. This information is shown on the following Excel spreadsheet. There is a slight difference in the software answer due to rounding. The Excel commands are similar to those detailed on page 219, number 2.



	A	B
1	Wearing Seat Belts	Probability
2	0	
3	1	
4	2	
5	3	
6	4	
7	5	
8	6	
9	7	0.0902
10	8	0.1805
11	9	0.2569
12	10	0.2467
13	11	0.1436
14	12	0.0383
15		0.9563
16		
17		

Excel

Self-Review 6–4 For a case where $n = 4$ and $\pi = .60$, determine the probability that:

(a) $x = 2$.
(b) $x \leq 2$.
(c) $x > 2$.

Exercises

19. In a binomial distribution $n = 8$ and $\pi = .30$. Find the probabilities of the following events.
   a. $x = 2$.
   b. $x \leq 2$ (the probability that x is equal to or less than 2).
   c. $x \geq 3$ (the probability that x is equal to or greater than 3).

20. In a binomial distribution $n = 12$ and $\pi = .60$. Find the following probabilities.
   a. $x = 5$.
   b. $x \leq 5$.
   c. $x \geq 6$.

21. In a recent study 90 percent of the homes in the United States were found to have large-screen TVs. In a sample of nine homes, what is the probability that:
   a. All nine have large-screen TVs?
   b. Less than five have large-screen TVs?
   c. More than five have large-screen TVs?
   d. At least seven homes have large-screen TVs?

22. A manufacturer of window frames knows from long experience that 5 percent of the production will have some type of minor defect that will require an adjustment. What is the probability that in a sample of 20 window frames:

a. None will need adjustment?
b. At least one will need adjustment?
c. More than two will need adjustment?

23. The speed with which utility companies can resolve problems is very important. GTC, the Georgetown Telephone Company, reports it can resolve customer problems the same day they are reported in 70 percent of the cases. Suppose the 15 cases reported today are representative of all complaints.
   a. How many of the problems would you expect to be resolved today? What is the standard deviation?
   b. What is the probability 10 of the problems can be resolved today?
   c. What is the probability 10 or 11 of the problems can be resolved today?
   d. What is the probability more than 10 of the problems can be resolved today?
24. Backyard Retreats, Inc., sells an exclusive line of pools, hot tubs, and spas. It is located just off the Bee Line Expressway in Orlando, Florida. The owner reports 20 percent of the customers entering the store will make a purchase of at least $50. Suppose 15 customers enter the store before 10 A.M. on a particular Saturday.
   a. How many of these customers would you expect to make a purchase of at least $50?
   b. What is the probability exactly five of these customers make a purchase of at least $50?
   c. What is the probability at least five of these customers make a purchase of at least $50?
   d. What is the probability at least one customer makes a purchase of at least $50?

Hypergeometric Probability Distribution

For the binomial distribution to be applied, the probability of a success must stay the same for each trial. For example, the probability of guessing the correct answer to a true/false question is .50. This probability remains the same for each question on an examination. Likewise, suppose that 40 percent of the registered voters in a precinct are Republicans. If 27 registered voters are selected at random, the probability of choosing a Republican on the first selection is .40. The chance of choosing a Republican on the next selection is also .40, assuming that the sampling is done *with replacement,* meaning that the person selected is put back in the population before the next person is selected.

Most sampling, however, is done *without replacement.* Thus, if the population is small, the probability for each observation will change. For example, if the population consists of 20 items, the probability of selecting a particular item from that population is 1/20. If the sampling is done without replacement, after the first selection there are only 19 items remaining; the probability of selecting a particular item on the second selection is only 1/19. For the third selection, the probability is 1/18, and so on. This assumes that the population is **finite**—that is, the number in the population is known and relatively small in number. Examples of a finite population are 2,842 Republicans in the precinct, 9,241 applications for medical school, and the 18 Pontiac Vibes currently in stock at North Charleston Pontiac.

Recall that one of the criteria for the binomial distribution is that the probability of success remains the same from trial to trial. Since the probability of success does not remain the same from trial to trial when sampling is from a relatively small population without replacement, the binomial distribution should not be used. Instead, the **hypergeometric distribution** is applied. Therefore, (1) if a sample is selected from a finite population without replacement and (2) if the size of the sample n is more than 5 percent of the size of the population $N,$ then the hypergeometric distribution is used to determine the probability of a specified number of successes or failures. It is especially appropriate when the size of the population is small.

The formula for the hypergeometric distribution is:

HYPERGEOMETRIC DISTRIBUTION $$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n}$$ **[6–6]**

where:

N is the size of the population.
S is the number of successes in the population.
x is the number of successes in the sample. It may be 0, 1, 2, 3, ...
n is the size of the sample or the number of trials.
C is the symbol for a combination.

In summary, a hypergeometric probability distribution has these characteristics:

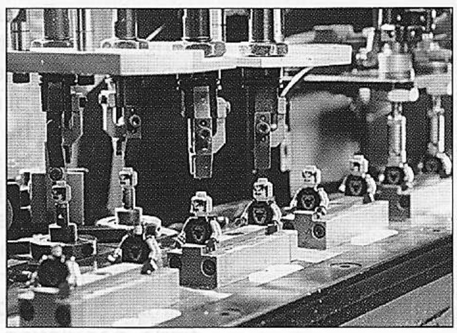
HYPERGEOMETRIC PROBABILITY EXPERIMENT

1. An outcome on each trial of an experiment is classified into one of two mutually exclusive categories—a success or a failure.
2. The random variable is the number of successes in a fixed number of trials.
3. The trials are *not independent*.
4. We assume that we sample from a finite population without replacement and $n/N > 0.05$. So, the probability of a success *changes* for each trial.

The following example illustrates the details of determining a probability using the hypergeometric distribution.

Example

PlayTime Toys, Inc., employs 50 people in the Assembly Department. Forty of the employees belong to a union and ten do not. Five employees are selected at random to form a committee to meet with management regarding shift starting times. What is the probability that four of the five selected for the committee belong to a union?



Solution

The population in this case is the 50 Assembly Department employees. An employee can be selected for the committee only once. Hence, the sampling is done without replacement. Thus, the probability of selecting a union employee, for example, changes from one trial to the next. The hypergeometric distribution is appropriate for determining the probability. In this problem,

N is 50, the number of employees.
S is 40, the number of union employees.
x is 4, the number of union employees selected.
n is 5, the number of employees selected.

We wish to find the probability 4 of the 5 committee members belong to a union. Inserting these values into formula (6–6):

$$P(4) = \frac{({}_{40}C_4)({}_{50-40}C_{5-4})}{{}_{50}C_5} = \frac{\left(\frac{40!}{4!36!}\right)\left(\frac{10!}{1!9!}\right)}{\frac{50!}{5!45!}} = \frac{(91{,}390)(10)}{2{,}118{,}760} = .431$$

Thus, the probability of selecting 5 assembly workers at random from the 50 workers and finding 4 of the 5 are union members is .431.

Table 6–4 shows the hypergeometric probabilities of finding 0, 1, 2, 3, 4, and 5 union members on the committee.

TABLE 6–4 Hypergeometric Probabilities ($n = 5$, $N = 50$, and $S = 40$) for the Number of Union Members on the Committee

Union Members	Probability
0	.000
1	.004
2	.044
3	.210
4	.431
5	.311
	1.000

In order for you to compare the two probability distributions, Table 6–5 shows the hypergeometric and binomial probabilities for the PlayTime Toys, Inc., example. Because 40 of the 50 Assembly Department employees belong to the union, we let $\pi = .80$ for the binomial distribution. The binomial probabilities for Table 6–5 come from the binomial distribution with $n = 5$ and $\pi = .80$.

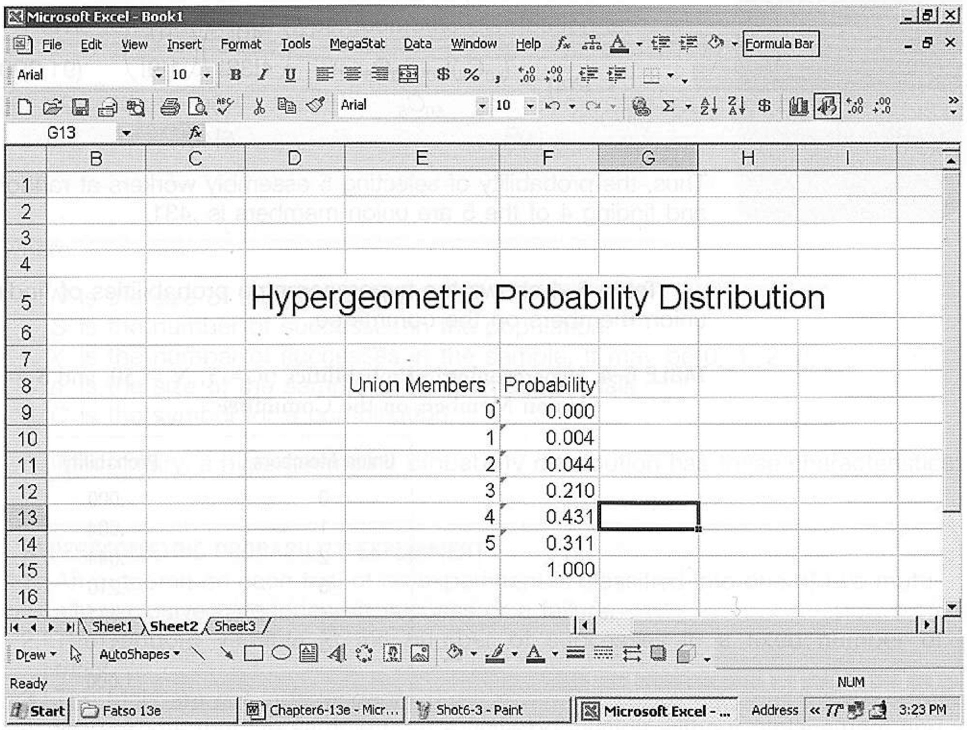
TABLE 6–5 Hypergeometric and Binomial Probabilities for PlayTime Toys, Inc., Assembly Department

Number of Union Members on Committee	Hypergeometric Probability, $P(x)$	Binomial Probability ($n = 5$ and $\pi = .80$)
0	.000	.000
1	.004	.006
2	.044	.051
3	.210	.205
4	.431	.410
5	.311	.328
	1.000	1.000

When the binomial requirement of a constant probability of success cannot be met, the hypergeometric distribution should be used. However, as Table 6–5 shows, under certain conditions the results of the binomial distribution can be used to approximate the hypergeometric. This leads to a rule of thumb:

If selected items are not returned to the population, the binomial distribution can be used to closely approximate the hypergeometric distribution when $n < .05N$. In words, the binomial will suffice if the sample size is less than 5 percent of the population.

A hypergeometric distribution can be created using Excel. See the following output. The necessary steps are given in the Software Commands section.



Union Members	Probability
0	0.000
1	0.004
2	0.044
3	0.210
4	0.431
5	0.311
	1.000

Self-Review 6–5

Horwege Discount Brokers plans to hire 5 new financial analysts this year. There is a pool of 12 approved applicants, and George Horwege, the owner, decides to randomly select those who will be hired. There are 8 men and 4 women among the approved applicants. What is the probability that 3 of the 5 hired are men?

Exercises

25. A population consists of 10 items, 6 of which are defective. In a sample of 3 items, what is the probability that exactly 2 are defective? Assume the samples are drawn without replacement.

26. A population consists of 15 items, 10 of which are acceptable. In a sample of 4 items, what is the probability that exactly 3 are acceptable? Assume the samples are drawn without replacement.

27. Kolzak Appliance Outlet just received a shipment of 10 DVD players. Shortly after they were received, the manufacturer called to report that he had inadvertently shipped 3 defective units. Ms. Kolzak, the owner of the outlet, decided to test 2 of the 10 DVD players she received. What is the probability that neither of the 2 DVD players tested is defective? Assume the samples are drawn without replacement.

28. The Computer Systems Department has eight faculty, six of whom are tenured. Dr. Vonder, the chairman, wants to establish a committee of three department faculty members to review the curriculum. If she selects the committee at random:

a. What is the probability all members of the committee are tenured?

b. What is the probability that at least one member is not tenured? (Hint: For this question, use the complement rule.)

29. Keith's Florists has 15 delivery trucks, used mainly to deliver flowers and flower arrangements in the Greenville, South Carolina, area. Of these 15 trucks, 6 have brake problems. A sample of 5 trucks is randomly selected. What is the probability that 2 of those tested have defective brakes?

30. The game called Lotto sponsored by the Louisiana Lottery Commission pays its largest prize when a contestant matches all 6 of the 40 possible numbers. Assume there are 40 ping-pong balls each with a single number between 1 and 40. Any number appears only once, and the winning balls are selected without replacement.

a. The commission reports that the probability of matching all the numbers are 1 in 3,838,380. What is this in terms of probability?

b. Use the hypergeometric formula to find this probability.

The lottery commission also pays if a contestant matches 4 or 5 of the 6 winning numbers. Hint: Divide the 40 numbers into two groups, winning numbers and nonwinning numbers.

c. Find the probability, again using the hypergeometric formula, for matching 4 of the 6 winning numbers.

d. Find the probability of matching 5 of the 6 winning numbers.

Poisson Probability Distribution

The **Poisson probability distribution** describes the number of times some event occurs during a specified interval. The interval may be time, distance, area, or volume.

The distribution is based on two assumptions. The first assumption is that the probability is proportional to the length of the interval. The second assumption is that the intervals are independent. To put it another way, the longer the interval, the larger the probability, and the number of occurrences in one interval does not affect the other intervals. This distribution is also a limiting form of the binomial distribution when the probability of a success is very small and n is large. It is often referred to as the "law of improbable events," meaning that the probability, π, of a particular event's happening is quite small. The Poisson distribution is a discrete probability distribution because it is formed by counting.

In summary, a Poisson probability distribution has these characteristics:

POISSON PROBABILITY EXPERIMENT

1. The random variable is the number of times some event occurs during a defined interval.
2. The probability of the event is proportional to the size of the interval.
3. The intervals do not overlap and are independent.

This distribution has many applications. It is used as a model to describe the distribution of errors in data entry, the number of scratches and other imperfections in newly painted car panels, the number of defective parts in outgoing shipments, the number of customers waiting to be served at a restaurant or waiting to get into an attraction at Disney World, and the number of accidents on I–75 during a three-month period.

The Poisson distribution can be described mathematically by the formula:

POISSON DISTRIBUTION $$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad [6\text{–}7]$$

Statistics in Action

Near the end of World War II, the Germans developed rocket bombs, which were fired at the city of London. The Allied military command didn't know whether these bombs were fired at random or whether they had an aiming device. To investigate, the city of London was divided into 586 square regions. The distribution of hits in each square was recorded as follows:

Hits	0	1	2	3	4	5
Regions	229	221	93	35	7	1

To interpret, the above chart indicates that 229 regions were not hit with one of the bombs. Seven regions were hit four *(continued)* times. Using the Poisson distribution, with a mean of 0.93 hits per region, the expected number of hits is as follows:

Hits	0	1	2	3	4	5 or more
Regions	231.2	215.0	100.0	31.0	7.2	1.6

Because the actual number of hits was close to the expected number of hits, the military command concluded that the bombs were falling at random. The Germans had not developed a bomb with an aiming device.

where:

μ (mu) is the mean number of occurrences (successes) in a particular interval.
e is the constant 2.71828 (base of the Napierian logarithmic system).
x is the number of occurrences (successes).
$P(x)$ is the probability for a specified value of x.

The mean number of successes, μ, can be determined by $n\pi$, where n is the total number of trials and π the probability of success.

MEAN OF A POISSON DISTRIBUTION $\mu = n\pi$ **[6–8]**

The variance of the Poisson is also equal to its mean. If, for example, the probability that a check cashed by a bank will bounce is .0003, and 10,000 checks are cashed, the mean and the variance for the number of bad checks is 3.0, found by $\mu = n\pi = 10{,}000(.0003) = 3.0$.

Recall that for a binomial distribution there is a fixed number of trials. For example, for a four-question multiple-choice test there can only be zero, one, two, three, or four successes (correct answers). The random variable, x, for a Poisson distribution, however, can assume an *infinite number of values*—that is, 0, 1, 2, 3, 4, 5, . . . However, *the probabilities become very small after the first few occurrences* (successes).

To illustrate the Poisson probability computation, assume baggage is rarely lost by Northwest Airlines. Most flights do not experience any mishandled bags; some have one bag lost; a few have two bags lost; rarely a flight will have three lost bags; and so on. Suppose a random sample of 1,000 flights shows a total of 300 bags were lost. Thus, the arithmetic mean number of lost bags per flight is 0.3, found by 300/1,000. If the number of lost bags per flight follows a Poisson distribution with $\mu = 0.3$, we can compute the various probabilities using formula (6–7):

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

For example, the probability of not losing any bags is:

$$P(0) = \frac{(0.3)^0(e^{-0.3})}{0!} = 0.7408$$

In other words, 74 percent of the flights will have no lost baggage. The probability of exactly one lost bag is:

$$P(1) = \frac{(0.3)^1(e^{-0.3})}{1!} = 0.2222$$

Thus, we would expect to find exactly one lost bag on 22 percent of the flights.

Poisson probabilities can also be found in the table in Appendix B.5.

Example

Recall from the previous illustration that the number of lost bags follows a Poisson distribution with a mean of 0.3. Use Appendix B.5 to find the probability that no bags will be lost on a particular flight. What is the probability exactly one bag will be lost on a particular flight? When should the supervisor become suspicious that a flight is having too many lost bags?

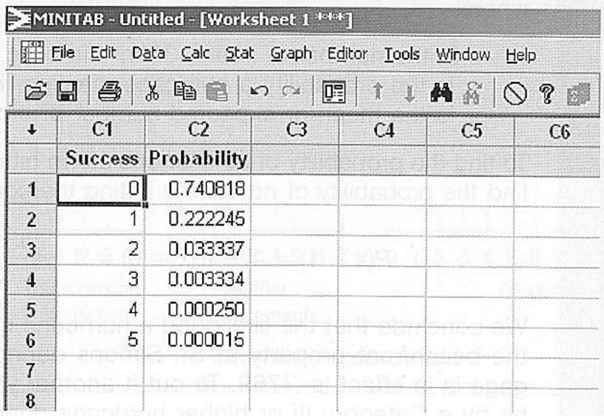
Solution

Part of Appendix B.5 is repeated as Table 6–6. To find the probability of no lost bags, locate the column headed "0.3" and read down that column to the row labeled "0." The probability is .7408. That is the probability of no lost bags. The probability of one lost bag is .2222, which is in the next row of the table, in the same column. The probability of two lost bags is .0333, in the row below; for three lost bags it is .0033; and for four lost bags it is .0003. Thus, a supervisor should not be surprised to find one lost bag but should expect to see more than one lost bag infrequently.

TABLE 6–6 Poisson Table for Various Values of μ (from Appendix B.5)

	μ								
x	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

These probabilities can also be found using the MINITAB system. The commands necessary are reported at the end of the chapter.



Earlier in this section we mentioned that the Poisson probability distribution is a limiting form of the binomial. That is, we could estimate a binomial probability using the Poisson.

The Poisson probability distribution is characterized by the number of times an event happens during some interval or continuum. Examples include:

- The number of misspelled words per page in a newspaper.
- The number of calls per hour received by Dyson Vacuum Cleaner Company.
- The number of vehicles sold per day at Hyatt Buick GMC in Durham, North Carolina.
- The number of goals scored in a college soccer game.

In each of these examples there is some type of continuum—misspelled words per page, calls per hour, vehicles per day, or goals per game.

In the previous example we investigated the number of bags lost per flight, so the continuum was a "flight." We knew the mean number of bags of luggage lost per flight, but we did not know the number of passengers or the probability of a bag being lost. We suspected the number of passengers was fairly large and the probability of a passenger losing his or her bag of luggage was small. In the following example we use the Poisson distribution to estimate a binomial probability when *n*, the number of trials, is large and π, the probability of a success, small.

Example

Coastal Insurance Company underwrites insurance for beachfront properties along the Virginia, North and South Carolina, and Georgia coasts. It uses the estimate that the probability of a named Category III hurricane (sustained winds of more than 110 miles per hour) or higher striking a particular region of the coast (for example, St. Simons Island, Georgia) in any one year is .05. If a homeowner takes a 30-year mortgage on a recently purchased property in St. Simons, what is the likelihood that the owner will experience at least one hurricane during the mortgage period?

Solution

To use the Poisson probability distribution we begin by determining the mean or expected number of storms meeting the criterion hitting St. Simons during the 30-year period. That is:

$$\mu = n\pi = 30(.05) = 1.5$$

where:

n is the number of years, 30 in this case.
π is the probability a hurricane meeting the strength criteria comes ashore.
μ is the mean or expected number of storms in a 30-year period.

To find the probability of at least one storm hitting St. Simons Island, Georgia, we first find the probability of no storms hitting the coast and subtract that value from 1.

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \frac{\mu^0 e^{-1.5}}{0!} = 1 - .2231 = .7769$$

We conclude that the likelihood a hurricane meeting the strength criteria will strike the beachfront property at St. Simons during the 30-year period when the mortgage is in effect is .7769. To put it another way, the probability St. Simons will be hit by a Category III or higher hurricane during the 30-year period is a little more than 75 percent.

We should emphasize that the continuum, as previously described, still exists. That is, there are expected to be 1.5 storms hitting the coast per 30-year period. The continuum is the 30-year period.

In the above case we are actually using the Poisson distribution as an estimate of the binomial. Note that we've met the binomial conditions outlined on page 190.

- There are only two possible outcomes: a hurricane hits the St. Simons area or it does not.
- There is a fixed number of trials, in this case 30 years.
- There is a constant probability of success; that is, the probability of a hurricane hitting the area is .05 each year.
- The years are independent. That means if a named storm strikes in the fifth year, that has no effect on any other year.

To find the probability of at least one storm striking the area in a 30-year period using the binomial distribution:

$$P(x \geq 1) = 1 - P(x = 0) = 1 - {}_{30}C_0(.05)^0(.95)^{30} = 1 - (1)(1)(.2146) = .7854$$

The probability of at least one hurricane hitting the St. Simons area during the 30-year period using the binomial distribution is .7854.

Which answer is correct? Why should we look at the problem both ways? The binomial is the more "technically correct" solution. The Poisson can be thought of as an approximation for the binomial, when n, the number of trials is large, and π, the probability of a success, is small. We look at the problem using both distributions to emphasize the convergence of the two discrete distributions. In some instances using the Poisson may be the quicker solution, and as you see there is little practical difference in the answers. In fact, as n gets larger and π smaller, the differences between the two distributions gets smaller.

The Poisson probability distribution is always positively skewed and the random variable has no specific upper limit. The Poisson distribution for the lost bags illustration, where $\mu = 0.3$, is highly skewed. As μ becomes larger, the Poisson distribution becomes more symmetrical. For example, Chart 6–4 shows the distributions of the number of transmission services, muffler replacements, and oil changes per day at Avellino's Auto Shop. They follow Poisson distributions with means of 0.7, 2.0, and 6.0, respectively.

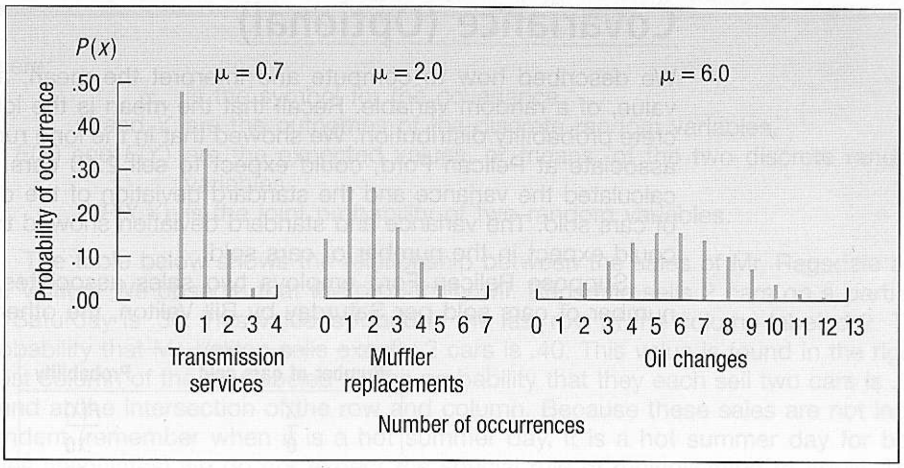


CHART 6–4 Poisson Probability Distributions for Means of 0.7, 2.0, and 6.0

Only μ needed to construct Poisson

In summary, the Poisson distribution is actually a family of discrete distributions. All that is needed to construct a Poisson probability distribution is the mean number of defects, errors, and so on—designated as μ.

Self-Review 6–6

From actuary tables Washington Insurance Company determined the likelihood that a man age 25 will die within the next year is .0002. If Washington Insurance sells 4,000 policies to 25-year-old men this year, what is the probability they will pay on exactly one policy?

Exercises

31. In a Poisson distribution $\mu = 0.4$.
a. What is the probability that $x = 0$?
b. What is the probability that $x > 0$?

32. In a Poisson distribution $\mu = 4$.
a. What is the probability that $x = 2$?
b. What is the probability that $x \leq 2$?
c. What is the probability that $x > 2$?

33. Ms. Bergen is a loan officer at Coast Bank and Trust. From her years of experience, she estimates that the probability is .025 that an applicant will not be able to repay his or her installment loan. Last month she made 40 loans.
a. What is the probability that 3 loans will be defaulted?
b. What is the probability that at least 3 loans will be defaulted?

34. Automobiles arrive at the Elkhart exit of the Indiana Toll Road at the rate of two per minute. The distribution of arrivals approximates a Poisson distribution.
a. What is the probability that no automobiles arrive in a particular minute?
b. What is the probability that at least one automobile arrives during a particular minute?

35. It is estimated that 0.5 percent of the callers to the Customer Service department of Dell, Inc., will receive a busy signal. What is the probability that of today's 1,200 callers at least 5 received a busy signal?

36. Textbook authors and publishers work very hard to minimize the number of errors in a text. However, some errors are unavoidable. Mr. J. A. Carmen, statistics editor, reports that the mean number of errors per chapter is 0.8. What is the probability that there are less than 2 errors in a particular chapter?

Covariance (Optional)

We described how to compute and interpret the mean, also called the expected value, of a random variable. Recall that the mean is the long run average of a discrete probability distribution. We showed that in the long run John Ragsdale, a sales associate at Pelican Ford, could expect to sell 2.10 cars each Saturday. Next we calculated the variance and the standard deviation of the distribution of the number of cars sold. The variance and standard deviation showed the variation Mr. Ragsdale could expect in the number of cars sold.

Suppose Pelican Ford employs two sales associates. The distribution of the number of cars sold per Saturday by Bill Valiton, the other associate, follows.

Number of cars sold X	Probability $P(X)$
0	.10
1	.50
2	.40

As a sales manager you are really interested in the *total* number of vehicles sold on a particular Saturday. That is, you are interested in the distribution of the total number of vehicles sold, rather than the individual distributions of Mr. Ragsdale and Mr. Valiton. We can find any linear combination of two random variables by the equation

THE LINEAR COMBINATION OF TWO RANDOM VARIABLES $Z = aX + bY$

where:

X and Y are the two random variables.
a and b are constants or weights.
Z is the sum of the products of two random variables.

If we wish to find the expected value of the sum of two random variables and $a = b = 1$, the above equation simplifies to $E(Z) = E(X) + E(Y)$. In words, this indicates that the mean of the distribution of the sum of two random variables is the sum of the two expected values or means.

In the Pelican Ford example, the mean number of vehicles sold by Mr. Valiton is 1.30, found by

$$\mu = E(Y) = \Sigma Y(P(Y)) = 0(.10) + 1(.50) + 2(.40) = 1.30$$

The mean, or expected value, of the total number of vehicles sold by the two associates is

$$E(Z) = E(X) + E(Y) = 2.10 + 1.30 = 3.40.$$

This presents a solution to only part of the problem. We can reason logically about what will happen to the mean, or the expected value, of the sum of two random variables. But we are also interested in the variation of the sum of these two variables. A confounding factor is the prospect that there is an interrelationship between the two variables. In the Pelican Ford example, it is reasonable that there is an interrelationship between the sales of Mr. Ragsdale and the sales of Mr. Valiton. For example, on a very hot summer Saturday prospective buyers are not going to stand out in the hot sun, so logically the sales for both associates is likely to be low.

The **covariance** is a measure of the relationship between two random variables.

THE COVARIANCE $\sigma_{xy} = \Sigma(X - E(X))(Y - E(Y))P(X, Y)$

where:

σ_{xy} is the symbol for the covariance.

X and Y are the outcomes of the discrete random variables.

$E(X)$ and $E(Y)$ are the expected values, or means, of the two discrete random variables.

$P(X,Y)$ is the joint probability of two random variables.

The table below shows the relationship between the sales of Mr. Ragsdale and Mr. Valiton. We observe that the probability Mr. Ragsdale sells 2 cars on a particular Saturday is .30. This value is found in the last row of the column labeled 2. The probability that Mr. Valiton sells exactly 2 cars is .40. This value is found in the rightmost column of the row labeled 2. The probability that they each sell two cars is .20, found at the intersection of the row and column. Because these sales are not independent (remember when it is a hot summer day, it is a hot summer day for both sales associates) we do not expect the special rule of multiplication to apply. That is $P(X,Y)$ does *not* equal $P(X)P(Y)$.

		Ragsdale Cars Sold (X)					
		0	**1**	**2**	**3**	**4**	**P(Y)**
	0	.05	.02	.03	.00	.00	.10
Valiton	1	.05	.15	.07	.20	.03	.50
Cars Sold (Y)	2	.00	.03	.20	.10	.07	.40
	P(X)	.10	.20	.30	.30	.10	1.00

To determine the covariance we use

$$\sigma_{xy} = \Sigma (X - E(X))(Y - E(Y))P(X, Y)$$

where

$$\sigma_{xy} = (0 - 2.1)(0 - 1.3).05 + (1 - 2.1)(0 - 1.3).02 + \cdots + (4 - 2.1)(2 - 1.3).07$$
$$= 0.95$$

The covariance reports how the two variables move together. The 0.95 indicates that the two variables are directly related. That is, when Mr. Ragsdale sells more than his mean number of cars, the tendency is for Mr. Valiton to sell more than his mean also.

The major drawback to the covariance is we have little understanding of the magnitude of the difference. The units are "cars squared." Is 0.9500 a lot or a little? We don't know. If the covariance were a negative value, this would indicate the two distributions were inversely or indirectly related. If it were 0, then the two distributions are unrelated *or independent.*

Because we now have information about the relationship between the two variables, we can reason about the variance of their sum. The variance of the sum of two random variables is found by

VARIANCE OF THE SUM OF TWO RANDOM VARIABLES $\sigma^2_{x+y} = a^2\sigma^2_x + b^2\sigma^2_y + 2ab\sigma_{xy}$

The values of *a* and *b,* as before, represent the weights or constants assigned. When $a = b = 1$, the equation simplifies to

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y + 2\sigma_{xy}$$

In words, the above equation indicates that the variance of the sum of two random variables is equal to the sum of the variances of the two random variables plus two times the covariance. This indicates that when we want to consider the sum of two variables, we need to account for the variation in each of the variables *plus* the interrelationship between the two variables.

To complete the question of the variability of the total number of cars sold on Saturday we need to find the variance of the distribution of the sales for Mr. Valiton. Using formula (6–2),

$$\sigma^2_y = \Sigma(Y - \mu)^2 P(Y) = (0 - 1.3)^2(.10) + (1 - 1.3)^2(.50) + (2 - 1.3)^2(.40) = 0.41$$

Recall that on page 187 we computed the variance of the distribution of the number of vehicles sold by Mr. Ragsdale to be 1.29. So the variance of the sum of the two random variables is

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y + 2\sigma_{XY} = 1.29 + 0.41 + 2(0.95) = 3.60$$

To summarize, the mean number of vehicles sold each Saturday at Pelican Ford is 3.40 vehicles and the variance is 3.60. The standard deviation is 1.8974 vehicles, found by taking the square root of 3.60.

One of the most useful applications of the above relationships is in the field of financial analysis. Investors are interested in obtaining the maximum rate of return, but they are also interested in reducing the risk. In statistical terms reducing the risk means reducing the variance or the standard deviation. The following example will help to explain the details.

Example

Ernie DuBrul just inherited $200,000 and will divide this inheritance into a portfolio of two investments. After some research Ernie decides to put 25 percent into American Funds World Cap and the remainder into Burger International Funds. For the American Funds World Cap the mean return is 12 percent and the standard deviation 3 percent. For the Burger International Funds the mean rate of return is

20 percent with a standard deviation of 8 percent. After some calculation, he is able to determine that the covariance between the two investments is 12. What is the expected rate of return for the portfolio? What should he conclude about the relationship between the two investments? What is the portfolio's standard deviation?

Solution

Ernie can consider the two investments as random variables with the means of 12 percent and 20 percent, respectively. The weight of the first investment is .25 (a = .25) and .75 (b = .75) for the second. The expected amount of return for the portfolio is 18 percent, found by

$$E(Z) = E(X + Y) = a(E(X)) + b(E(Y)) = .25(12) + .75(20) = 18.0$$

The covariance of 12 suggests that there is a positive relationship between the two investments, because it is a positive number. However, the value of 12 does not give much insight into the strength of the relationship.

We can determine the variance of the portfolio as follows.

$$\sigma^2_{x+y} = a^2\sigma^2_x + b^2\sigma^2_y + 2ab\sigma_{xy} = (.25)^2(3)^2 + (.75)^2(8)^2 + 2(.25)(.75)(12) = 41.0625$$

The square root of 41.0625 is 6.4 percent, which is the standard deviation of the weighted sum of the two variables.

How does Ernie interpret this information? Suppose he had an opportunity to invest all $200,000 in Internet stocks, where the rate of return was the same, 18 percent, but the standard deviation of this distribution was 8.0 percent. The standard deviation of 8.0 percent indicates there is more risk in the investment of Internet stocks. Most investors want to reduce their risk; hence, the better course of action is to make the investment he originally planned.

In the above examples there was association between the two distributions; that is, the covariance was not equal to 0. Consider the following example where there is no association between the two distributions.

Example

Suppose we are involved in a game with 2 fair coins. The coins are flipped and we count the number of heads. For each head that appears we are given $1.00 by the House; for each tail we must pay the House the same amount. We can summarize the game with the following table.

		Coin 1		
		Head $1	Tail −$1	Total
Coin 2	Head $1	.25	.25	.50
	Tail $−1	.25	.25	.50
	Total	.50	.50	1.00

The means of the two random variables are:

$$E(X) = \$1(.50) + (\$ - 1)(.50) = \$0.00$$

$$E(Y) = \$1(.50) + (\$ - 1)(.50) = \$0.00$$

The variances of the two random variables are:

$$\sigma^2_x = (1 - 0)^2.50 + (-1 - 0)^2(.50) = 1$$

$$\sigma^2_y = (1 - 0)^2.50 + (-1 - 0)^2(.50) = 1$$

The covariance of the two random variables is:

$$\sigma_{xy} = (1-0)(1-0).25 + (-1-0)(1-0).25 + (-1-0)(1-0).25 + (-1-0)(-1-0).25$$

$$\sigma_{xy} = (1).25 + (-1).25 + (-1).25 + (1).25 = 0$$

The fact that the covariance is 0 indicates that the two variables are not related, that is, they are independent. That is, the outcome of the first coin is not related to the outcome of the second coin. We already knew this from our earlier study of probability, but the fact that the covariance is 0 confirms it.

Exercises

O–1. Two random variables are given by the following table.

	0	1	2	$P(y)$
0	.3	.1	0	.4
1	.1	.3	.1	.5
2	0	0	.1	.1
$P(X)$	.4	.4	.2	1.00

a. Find the mean of the variables *x* and *y*.
b. Find the variance of the variables *x* and *y*.
c. Find the covariance.
d. Find the expected value of the sum of the two variables.
e. Find the variance of the sum of the two variables.

O–2. An analysis of two stocks indicates that the mean rate of return of the first is 8 percent with a standard deviation of 15 percent. The second has a mean rate of return of 14 percent with a standard deviation of 20 percent. Suppose we invest 40 percent in the first stock and 60 percent in the second.
a. What is the expected rate of return on the total investment?
b. If the two stocks are not related, what is the standard deviation of the rate of return on the total investment?
c. Suppose the covariance between the two stocks is 150. What is the standard deviation for the rate of return?

Chapter Summary

I. A random variable is a numerical value determined by the outcome of an experiment.
II. A probability distribution is a listing of all possible outcomes of an experiment and the probability associated with each outcome.
A. A discrete probability distribution can assume only certain values. The main features are:
1. The sum of the probabilities is 1.00.
2. The probability of a particular outcome is between 0.00 and 1.00.
3. The outcomes are mutually exclusive.
B. A continuous distribution can assume an infinite number of values within a specific range.
III. The mean and variance of a probability distribution are computed as follows.
A. The mean is equal to:

$$\mu = \Sigma[xP(x)] \qquad \text{[6–1]}$$

B. The variance is equal to:

$$\sigma^2 = \Sigma[(x - \mu)^2 P(x)] \quad \textbf{[6–2]}$$

IV. The binomial distribution has the following characteristics.

A. Each outcome is classified into one of two mutually exclusive categories.

B. The distribution results from a count of the number of successes in a fixed number of trials.

C. The probability of a success remains the same from trial to trial.

D. Each trial is independent.

E. A binomial probability is determined as follows:

$$P(x) = {}_nC_x \pi^x (1 - \pi)^{n-x} \quad \textbf{[6–3]}$$

F. The mean is computed as:

$$\mu = n\pi \quad \textbf{[6–4]}$$

G. The variance is

$$\sigma^2 = n\pi(1 - \pi) \quad \textbf{[6–5]}$$

V. The hypergeometric distribution has the following characteristics.

A. There are only two possible outcomes.

B. The probability of a success is not the same on each trial.

C. The distribution results from a count of the number of successes in a fixed number of trials.

D. It is used when sampling without replacement from a finite population.

E. A hypergeometric probability is computed from the following equation:

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{({}_NC_n)} \quad \textbf{[6–6]}$$

VI. The Poisson distribution has the following characteristics.

A. It describes the number of times some event occurs during a specified interval.

B. The probability of a "success" is proportional to the length of the interval.

C. Nonoverlapping intervals are independent.

D. It is a limiting form of the binomial distribution when n is large and π is small.

E. A Poisson probability is determined from the following equation:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \textbf{[6–7]}$$

F. The mean and the variance are:

$$\mu = n\pi \quad \textbf{[6–8]}$$

$$\sigma^2 = n\pi$$

Chapter Exercises

37. What is the difference between a random variable and a probability distribution?

38. For each of the following indicate whether the random variable is discrete or continuous.

a. The length of time to get a haircut.

b. The number of cars a jogger passes each morning while running.

c. The number of hits for a team in a high school girls' softball game.

d. The number of patients treated at the South Strand Medical Center between 6 and 10 P.M. each night.

e. The distance your car traveled on the last fill-up.

f. The number of customers at the Oak Street Wendy's who used the drive-through facility.

g. The distance between Gainesville, Florida, and all Florida cities with a population of at least 50,000.

39. What are the requirements for the binomial distribution?

40. Under what conditions will the binomial and the Poisson distributions give roughly the same results?

41. Samson Apartments, Inc., has a large number of units. A concern of management is the number of vacant apartments. A recent study revealed the percent of the time that a given number of apartments are vacant. Compute the mean and standard deviation of the number of vacant apartments.

Number of Vacant Units	Probability
0	.1
1	.2
2	.3
3	.4

42. An investment will be worth $1,000, $2,000, or $5,000 at the end of the year. The probabilities of these values are .25, .60, and .15, respectively. Determine the mean and variance of the worth of the investment.

43. The personnel manager of Cumberland Pig Iron Company is studying the number of on-the-job accidents over a period of one month. He developed the following probability distribution. Compute the mean, variance, and standard deviation of the number of accidents in a month.

Number of Accidents	Probability
0	.40
1	.20
2	.20
3	.10
4	.10

44. Croissant Bakery, Inc., offers special decorated cakes for birthdays, weddings, and other occasions. It also has regular cakes available in its bakery. The following table gives the total number of cakes sold per day and the corresponding probability. Compute the mean, variance, and standard deviation of the number of cakes sold per day.

Number of Cakes Sold in a Day	Probability
12	.25
13	.40
14	.25
15	.10

45. A Tamiami shearing machine is producing 10 percent defective pieces, which is abnormally high. The quality control engineer has been checking the output by almost continuous sampling since the abnormal condition began. What is the probability that in a sample of 10 pieces:

a. Exactly 5 will be defective?

b. 5 or more will be defective?

46. Thirty percent of the population in a southwestern community are Spanish-speaking Americans. A Spanish-speaking person is accused of killing a non-Spanish-speaking American and goes to trial. Of the first 12 potential jurors, only 2 are Spanish-speaking Americans, and 10 are not. The defendant's lawyer challenges the jury selection, claiming bias against her client. The government lawyer disagrees, saying that the probability of this particular jury composition is common. Compute the probability and discuss the assumptions.

47. An auditor for Health Maintenance Services of Georgia reports 40 percent of policyholders 55 years or older submit a claim during the year. Fifteen policyholders are randomly selected for company records.

a. How many of the policyholders would you expect to have filed a claim within the last year?
b. What is the probability that 10 of the selected policyholders submitted a claim last year?
c. What is the probability that 10 or more of the selected policyholders submitted a claim last year?
d. What is the probability that more than 10 of the selected policyholders submitted a claim last year?

48. Tire and Auto Supply is considering a 2-for-1 stock split. Before the transaction is finalized, at least two-thirds of the 1,200 company stockholders must approve the proposal. To evaluate the likelihood the proposal will be approved, the CFO selected a sample of 18 stockholders. He contacted each and found 14 approved of the proposed split. What is the likelihood of this event, assuming two-thirds of the stockholders approve?

49. A federal study reported that 7.5 percent of the U.S. workforce has a drug problem. A drug enforcement official for the State of Indiana wished to investigate this statement. In her sample of 20 employed workers:
a. How many would you expect to have a drug problem? What is the standard deviation?
b. What is the likelihood that *none* of the workers sampled has a drug problem?
c. What is the likelihood *at least one* has a drug problem?

50. The Bank of Hawaii reports that 7 percent of its credit card holders will default at some time in their life. The Hilo branch just mailed out 12 new cards today.
a. How many of these new cardholders would you expect to default? What is the standard deviation?
b. What is the likelihood that *none* of the cardholders will default?
c. What is the likelihood *at least one* will default?

51. Recent statistics suggest that 15 percent of those who visit a retail site on the World Wide Web make a purchase. A retailer wished to verify this claim. To do so, she selected a sample of 16 "hits" to her site and found that 4 had actually made a purchase.
a. What is the likelihood of exactly four purchases?
b. How many purchases should she expect?
c. What is the likelihood that four or more "hits" result in a purchase?

52. In Chapter 19 we discuss the *acceptance sample.* Acceptance sampling is used to monitor the quality of incoming raw materials. Suppose a purchaser of electronic components allows 1 percent of the components to be defective. To ensure the quality of incoming parts, a purchaser or manufacturer normally samples 20 parts and allows 1 defect.
a. What is the likelihood of accepting a lot that is 1 percent defective?
b. If the quality of the incoming lot was actually 2 percent, what is the likelihood of accepting it?
c. If the quality of the incoming lot was actually 5 percent, what is the likelihood of accepting it?

53. Colgate-Palmolive, Inc., recently developed a new toothpaste flavored with honey. It tested a group of ten people. Six of the group said they liked the new flavor, and the remaining four indicated they definitely did not. Four of the ten are selected to participate in an in-depth interview. What is the probability that of those selected for the in-depth interview two liked the new flavor and two did not?

54. Dr. Richmond, a psychologist, is studying the daytime television viewing habits of college students. She believes 45 percent of college students watch soap operas during the afternoon. To further investigate, she selects a sample of 10.
a. Develop a probability distribution for the number of students in the sample who watch soap operas.
b. Find the mean and the standard deviation of this distribution.
c. What is the probability of finding exactly four watch soap operas?
d. What is the probability less than half of the students selected watch soap operas?

55. A recent study conducted by Penn, Shone, and Borland, on behalf of LastMinute.com, revealed that 52 percent of business travelers plan their trips less than two weeks before departure. The study is to be replicated in the tri-state area with a sample of 12 frequent business travelers.
a. Develop a probability distribution for the number of travelers who plan their trips within two weeks of departure.
b. Find the mean and the standard deviation of this distribution.

c. What is the probability exactly 5 of the 12 selected business travelers plan their trips within two weeks of departure?
d. What is the probability 5 or fewer of the 12 selected business travelers plan their trips within two weeks of departure?

56. Suppose the Internal Revenue Service is studying the category of charitable contributions. A sample of 25 returns is selected from young couples between the ages of 20 and 35 who had an adjusted gross income of more than $100,000. Of these 25 returns five had charitable contributions of more than $1,000. Suppose four of these returns are selected for a comprehensive audit.
a. Explain why the hypergeometric distribution is appropriate.
b. What is the probability exactly one of the four audited had a charitable deduction of more than $1,000?
c. What is the probability at least one of the audited returns had a charitable contribution of more than $1,000?

57. The law firm of Hagel and Hagel is located in downtown Cincinnati. There are 10 partners in the firm; 7 live in Ohio and 3 in northern Kentucky. Ms. Wendy Hagel, the managing partner, wants to appoint a committee of 3 partners to look into moving the firm to northern Kentucky. If the committee is selected at random from the 10 partners, what is the probability that:
a. One member of the committee lives in northern Kentucky and the others live in Ohio?
b. At least 1 member of the committee lives in northern Kentucky?

58. Recent information published by the U.S. Environmental Protection Agency indicates that Honda is the manufacturer of four of the top nine vehicles in terms of fuel economy.
a. Determine the probability distribution for the number of Hondas in a sample of three cars chosen from the top nine.
b. What is the likelihood that in the sample of three at least one Honda is included?

59. The position of chief of police in the city of Corry, Pennsylvania, is vacant. A search committee of Corry residents is charged with the responsibility of recommending a new chief to the city council. There are 12 applicants, 4 of which are either female or members of a minority. The search committee decides to interview all 12 of the applicants. To begin, they randomly select four applicants to be interviewed on the first day, and none of the four is female or a member of a minority. The local newspaper, the Corry *Press,* suggests discrimination in an editorial. What is the likelihood of this occurrence?

60. Listed below is the population, according to 2004 estimates, by state for the 15 states with the largest population. Also included is whether that state's border touches the Gulf of Mexico, the Atlantic Ocean, or the Pacific Ocean (coastline).

Rank	State	Population	Coastline
1	California	35,893,799	Yes
2	Texas	22,490,022	Yes
3	New York	19,227,088	Yes
4	Florida	17,397,161	Yes
5	Illinois	12,713,634	No
6	Pennsylvania	12,406,292	No
7	Ohio	11,459,011	No
8	Michigan	10,112,620	No
9	Georgia	8,829,383	Yes
10	New Jersey	8,698,879	Yes
11	North Carolina	8,541,221	Yes
12	Virginia	7,459,827	Yes
13	Massachusetts	6,416,505	Yes
14	Indiana	6,237,569	No
15	Washington	6,203,788	Yes

Note that 5 of the 15 states do not have any coastline. Suppose three states are selected at random. What is the probability that:
a. None of the states selected have any coastline?
b. Exactly one of the selected states has a coastline.
c. At least one of the selected states has a coastline.

61. The sales of Lexus automobiles in the Detroit area follow a Poisson distribution with a mean of 3 per day.
a. What is the probability that no Lexus is sold on a particular day?
b. What is the probability that for five consecutive days at least one Lexus is sold?

62. Suppose 1.5 percent of the antennas on new Nokia cell phones are defective. For a random sample of 200 antennas, find the probability that:
a. None of the antennas is defective.
b. Three or more of the antennas are defective.

63. A study of the checkout lines at the Safeway Supermarket in the South Strand area revealed that between 4 and 7 P.M. on weekdays there is an average of four customers waiting in line. What is the probability that you visit Safeway today during this period and find:
a. No customers are waiting?
b. Four customers are waiting?
c. Four or fewer are waiting?
d. Four or more are waiting?

64. An internal study by the Technology Services department at Lahey Electronics revealed company employees receive an average of two emails per hour. Assume the arrival of these emails is approximated by the Poisson distribution.
a. What is the probability Linda Lahey, company president, received exactly 1 email between 4 P.M. and 5 P.M. yesterday?
b. What is the probability she received 5 or more email during the same period?
c. What is the probability she did not receive any email during the period?

65. Recent crime reports indicate that 3.1 motor vehicle thefts occur each minute in the United States. Assume that the distribution of thefts per minute can be approximated by the Poisson probability distribution.
a. Calculate the probability exactly *four* thefts occur in a minute.
b. What is the probability there are *no* thefts in a minute?
c. What is the probability there is *at least one* theft in a minute?

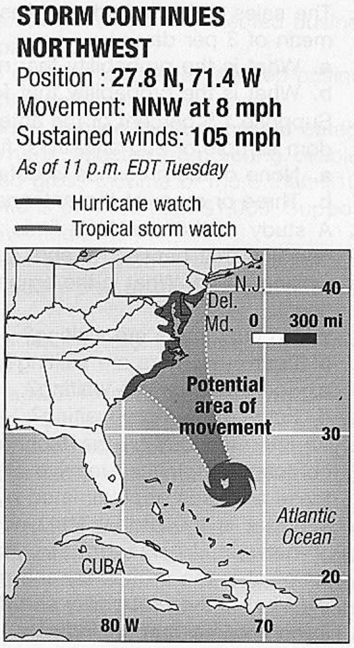
66. New Process, Inc., a large mail-order supplier of women's fashions, advertises same-day service on every order. Recently the movement of orders has not gone as planned, and there were a large number of complaints. Bud Owens, director of customer service, has completely redone the method of order handling. The goal is to have fewer than five unfilled orders on hand at the end of 95 percent of the working days. Frequent checks of the unfilled orders at the end of the day reveal that the distribution of the unfilled orders follows a Poisson distribution with a mean of two orders.
a. Has New Process, Inc., lived up to its internal goal? Cite evidence.
b. Draw a histogram representing the Poisson probability distribution of unfilled orders.

67. The National Aeronautics and Space Administration (NASA) has experienced two disasters. The Challenger exploded over the Atlantic Ocean in 1986 and the Columbia exploded over East Texas in 2003. There have been a total of 113 space missions. Assume failures continue to occur at the same rate and consider the next 23 missions. What is the probability of exactly two failures? What is the probability of no failures?

68. According to the "January theory," if the stock market is up for the month of January, it will be up for the year. If it is down in January, it will be down for the year. According to an article in *The Wall Street Journal,* this theory held for 29 out of the last 34 years. Suppose there is no truth to this theory; that is, the probability it is either up or down is .50. What is the probability this could occur by chance? (You will probably need a software package such as Excel or MINITAB.)

69. During the second round of the 1989 U.S. Open golf tournament, four golfers scored a hole in one on the sixth hole. The odds of a professional golfer making a hole in one are estimated to be 3,708 to 1, so the probability is 1/3,709. There were 155 golfers participating in the second round that day. Estimate the probability that four golfers would score a hole in one on the sixth hole.

70. On September 18, 2003, hurricane Isabel struck the North Carolina Coast, causing extensive damage. For several days prior to reaching land the National Hurricane Center had been predicting the hurricane would come on shore between Cape Fear, North Carolina, and the North Carolina–Virginia border. It was estimated that the probability the hurricane would actually strike in this area was .95. In fact, the hurricane did come on shore almost exactly as forecast and was almost in the center of the strike area.



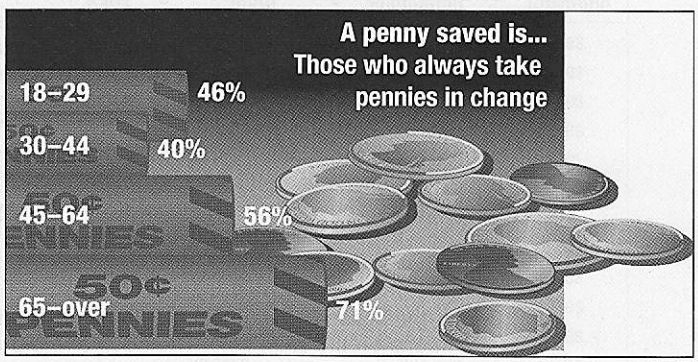
Suppose the National Hurricane Center forecasts that hurricanes will hit the strike area with a .95 probability. Answer the following questions:

a. What probability distribution does this follow?

b. What is the probability that 10 hurricanes reach landfall in the strike area?

c. What is the probability at least one of 10 hurricanes reaches land outside the strike area?

71. A recent CBS News survey reported that 67 percent of adults felt the U.S. Treasury should continue making pennies.



Suppose we select a sample of 15 adults.

a. How many of the 15 would we expect to indicate that the Treasury should continue making pennies? What is the standard deviation?

b. What is the likelihood that exactly 8 adults would indicate the Treasury should continue making pennies?

c. What is the likelihood at least 8 adults would indicate the Treasury should continue making pennies?

Data Set Exercises

72. Refer to the Real Estate data, which report information on homes sold in the Denver, Colorado, area last year.

a. Create a probability distribution for the number of bedrooms. Compute the mean and the standard deviation of this distribution.

b. Create a probability distribution for the number of bathrooms. Compute the mean and the standard deviation of this distribution.

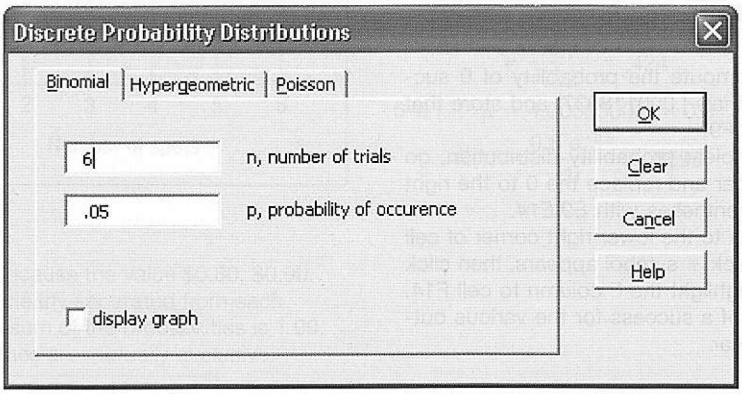
73. Refer to the Global Financial Performance data set that reports information on 148 corporations. There are 26 firms in the Financial Economic Sector. The corporate executive officers (CEOs) would like to meet to discuss the future of global financial services. They would like to randomly select 4 firms to participate. What is the probability that all four firms are from the commercial banking industry? Suggest a way to sample from the corporations so that one firm from each industry is included in the group.

Software Commands

1. The MegaStat commands to create the binomial probability distribution on page 193 are:

a. Select the **MegaStat** option on the toolbar, click on **Probability,** and **Discrete Probability Distributions.**

b. In the dialog box select the **Binomial** tab. The number of trials is *6*, the probability of a success is *.05*. If you wish to see a graph, click on **display graph.**



2. The Excel commands necessary to determine the binomial probability distribution on page 194 are:

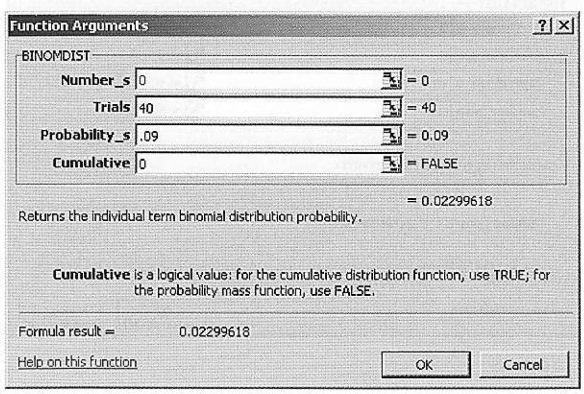
a. On a blank Excel worksheet write the word *Success* in cell A1 and the word *Probability* in B1. In cells A2 through A17 write the integers *0* to *15*. Click on *B2* as the active cell.

b. From the toolbar select **Insert,** then **Function.**

c. In the first dialog box select **Statistical** in the function category and **BINOMDIST** in the function name category, then click **OK.**

d. In the second dialog box enter the four items necessary to compute a binomial probability.

1. Enter *0* for the number of successes.
2. Enter *40* for the number of trials.



3. Enter *.09* for the probability of a success.
4. Enter the word *false* or the number *0* for the individual probabilities and click on **OK.**
5. Excel will compute the probability of 0 successes in 40 trials, with a .09 probability of success. The result, .02299618, is stored in cell B2.

e. To find the complete probability distribution, go to the formula bar and replace the *0* to the right of the open parentheses with *A2:A17.*

f. Move the mouse to the lower right corner of cell B2 till a solid black + symbol appears, then click and hold and highlight the B column to cell B17. The probability of a success for the various values of the random variable will appear.

3. The Excel commands necessary to determine the hypergeometric distribution on page 202 are:

a. On a blank Excel worksheet write the words *Union Members* in cell E8 and the word *Probability* in F8. In cells E9 to E14 write the integers *0* to *5*. Click on F9 as the active cell.

b. From the toolbar choose **Insert** and **Function.**

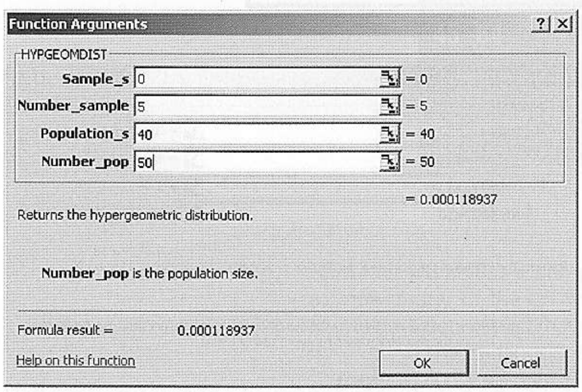
c. In the first dialog box select **Statistical** and **HYPGEOMDIST,** and then click **OK.**

d. In the second dialog box enter the four items necessary to compute a hypergeometric probability.

1. Enter *0* for the number of successes.
2. Enter *5* for the number of trials.
3. Enter *40* for the number of successes in the population.
4. Enter *50* for the size of the population and click **OK.**
5. Excel will compute the probability of 0 successes in 5 trials (.000118937) and store that result in cell F9.

e. To find the complete probability distribution, go to the formula bar and replace the 0 to the right of the open parentheses with *E9:E14.*

f. Move the mouse to the lower right corner of cell F9 till a solid black + symbol appears, then click and hold and highlight the F column to cell F14. The probability of a success for the various outcomes will appear.

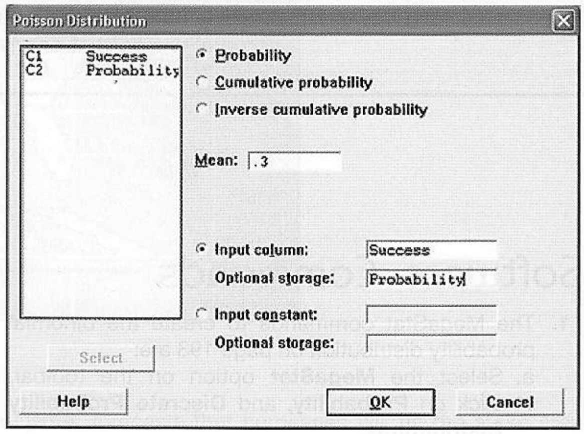


4. The MINITAB commands to generate the Poisson distribution on page 205 are:

a. Label column C1 as *Successes* and C2 as *Probability.* Enter the integers 0 though 5 in the first column.

b. Select **Calc,** then **Probability Distributions,** and **Poisson.**

c. In the dialog box click on **Probability,** set the mean equal to *.3,* and select *C1* as the Input column. Designate *C2* as Optional storage, and then click **OK.**

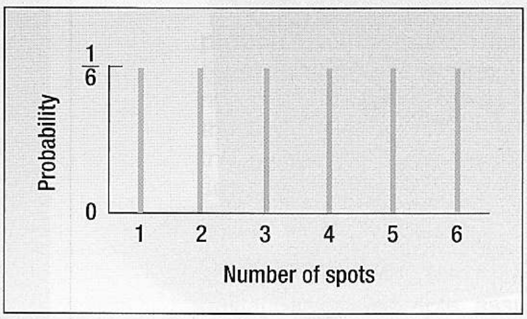


Chapter 6 Answers to Self-Review

6–1 a.

Number of Spots	Probability
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$
Total	$\frac{6}{6} = 1.00$

b.



c. $\frac{6}{6}$ or 1.

6–2 a. It is discrete, because the value \$0.80, \$0.90, and \$1.20 are clearly separated from each other. Also the sum of the probabilities is 1.00, and the outcomes are mutually exclusive.

b.

x	$P(x)$	$xP(x)$
\$ .80	.30	0.24
.90	.50	0.45
1.20	.20	0.24
		0.93

The mean is 93 cents.

c.

x	$P(x)$	$(x - \mu)$	$(x - \mu)^2 P(x)$
\$0.80	.30	−0.13	.00507
0.90	.50	−0.03	.00045
1.20	.20	0.27	.01458
			.02010

The variance is .02010, and the standard deviation is 14 cents.

6–3 a. It is reasonable because each employee either uses direct deposit or does not; employees are independent; the probability of using direct deposit is .80 for all; and we count the number using the service out of 7.

b. $P(7) = {}_7C_7\,(.80)^7\,(.20)^0 = .2097$

c. $P(4) = {}_7C_4\,(.80)^4\,(.20)^3 = .1147$

d. Answers are in agreement.

6–4 $n = 4, \pi = .60$

a. $P(x = 2) = .346$

b. $P(x \leq 2) = .526$

c. $P(x > 2) = 1 - .526 = .474$

6–5

$$P(3) = \frac{{}_8C_3\,{}_4C_2}{{}_{12}C_5} = \frac{\left(\frac{8!}{3!5!}\right)\left(\frac{4!}{2!2!}\right)}{\frac{12!}{5!7!}}$$

$$= \frac{(56)(6)}{792} = .424$$

6–6

$$\mu = 4{,}000(.0002) = 0.8$$

$$P(1) = \frac{0.8^1 e^{-0.8}}{1!} = .3595$$