CHAPTER 5

Discrete Probability Distributions

Objectives

After completing this chapter, you should be able to

1. Construct a probability distribution for a random variable.
2. Find the mean, variance, standard deviation, and expected value for a discrete random variable.
3. Find the exact probability for X successes in n trials of a binomial experiment.
4. Find the mean, variance, and standard deviation for the variable of a binomial distribution.
5. Find probabilities for outcomes of variables, using the Poisson, hypergeometric, and multinomial distributions.

Outline

Statistics Today

Is Pooling Worthwhile?

Blood samples are used to screen people for certain diseases. When the disease is rare, health care workers sometimes combine or pool the blood samples of a group of individuals into one batch and then test it. If the test result of the batch is negative, no further testing is needed since none of the individuals in the group has the disease. However, if the test result of the batch is positive, each individual in the group must be tested.

Consider this hypothetical example: Suppose the probability of a person having the disease is 0.05, and a pooled sample of 15 individuals is tested. What is the probability that no further testing will be needed for the individuals in the sample? The answer to this question can be found by using what is called the *binomial distribution*. See Statistics Today—Revisited at the end of the chapter.

This chapter explains probability distributions in general and a specific, often used distribution called the binomial distribution. The Poisson, hypergeometric, and multinomial distributions are also explained.

5–1 Introduction

Many decisions in business, insurance, and other real-life situations are made by assigning probabilities to all possible outcomes pertaining to the situation and then evaluating the results. For example, a saleswoman can compute the probability that she will make 0, 1, 2, or 3 or more sales in a single day. An insurance company might be able to assign probabilities to the number of vehicles a family owns. A self-employed speaker might be able to compute the probabilities for giving 0, 1, 2, 3, or 4 or more speeches each week. Once these probabilities are assigned, statistics such as the mean, variance, and standard deviation can be computed for these events. With these statistics, various decisions can be made. The saleswoman will be able to compute the average number of sales she makes per week, and if she is working on commission, she will be able to approximate her weekly income over a period of time, say, monthly. The public speaker will be able to plan ahead and approximate his average income and expenses. The insurance company can use its information to design special computer forms and programs to accommodate its customers' future needs.

This chapter explains the concepts and applications of what is called a *probability distribution.* In addition, special probability distributions, such as the *binomial, multinomial, Poisson,* and *hypergeometric* distributions, are explained.

5–2 Probability Distributions

Objective 1

Construct a probability distribution for a random variable.

Before probability distribution is defined formally, the definition of a variable is reviewed. In Chapter 1, a *variable* was defined as a characteristic or attribute that can assume different values. Various letters of the alphabet, such as *X, Y,* or *Z,* are used to represent variables. Since the variables in this chapter are associated with probability, they are called *random variables.*

For example, if a die is rolled, a letter such as *X* can be used to represent the outcomes. Then the value that *X* can assume is 1, 2, 3, 4, 5, or 6, corresponding to the outcomes of rolling a single die. If two coins are tossed, a letter, say *Y,* can be used to represent the number of heads, in this case 0, 1, or 2. As another example, if the temperature at 8:00 A.M. is 43° and at noon it is 53°, then the values *T* that the temperature assumes are said to be random, since they are due to various atmospheric conditions at the time the temperature was taken.

A **random variable** is a variable whose values are determined by chance.

Also recall from Chapter 1 that one can classify variables as discrete or continuous by observing the values the variable can assume. If a variable can assume only a specific number of values, such as the outcomes for the roll of a die or the outcomes for the toss of a coin, then the variable is called a *discrete variable.*

Discrete variables have a finite number of possible values or an infinite number of values that can be counted. The word *counted* means that they can be enumerated using the numbers 1, 2, 3, etc. For example, the number of joggers in Riverview Park each day and the number of phone calls received after a TV commercial airs are examples of discrete variables, since they can be counted.

Variables that can assume all values in the interval between any two given values are called *continuous variables.* For example, if the temperature goes from 62 to 78° in a 24-hour period, it has passed through every possible number from 62 to 78. *Continuous random variables are obtained from data that can be measured rather than counted.* Continuous random variables can assume an infinite number of values and can be decimal and fractional values. On a continuous scale, a person's weight might be exactly 183.426 pounds if a scale could measure weight to the thousandths place; however, on a digital scale that measures only to tenths of pounds, the weight would be 183.4 pounds. Examples of continuous variables are heights, weights, temperatures, and time. In this chapter only discrete random variables are used; Chapter 6 explains continuous random variables.

The procedure shown here for constructing a probability distribution for a discrete random variable uses the probability experiment of tossing three coins. Recall that when three coins are tossed, the sample space is represented as TTT, TTH, THT, HTT, HHT, HTH, THH, HHH; and if *X* is the random variable for the number of heads, then *X* assumes the value 0, 1, 2, or 3.

Probabilities for the values of X can be determined as follows:

No heads	One head			Two heads			Three heads
TTT	TTH	THT	HTT	HHT	HTH	THH	HHH
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$\frac{1}{8}$	$\frac{3}{8}$			$\frac{3}{8}$			$\frac{1}{8}$

Hence, the probability of getting no heads is $\frac{1}{8}$, one head is $\frac{3}{8}$, two heads is $\frac{3}{8}$, and three heads is $\frac{1}{8}$. From these values, a probability distribution can be constructed by listing the outcomes and assigning the probability of each outcome, as shown here.

Number of heads X	0	1	2	3
Probability $P(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A **discrete probability distribution** consists of the values a random variable can assume and the corresponding probabilities of the values. The probabilities are determined theoretically or by observation.

Discrete probability distributions can be shown by using a graph or a table. Probability distributions can also be represented by a formula. See Exercises 31–36 at the end of this section for examples.

Example 5–1

Construct a probability distribution for rolling a single die.

Solution

Since the sample space is 1, 2, 3, 4, 5, 6 and each outcome has a probability of $\frac{1}{6}$, the distribution is as shown.

Outcome X	1	2	3	4	5	6
Probability $P(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Probability distributions can be shown graphically by representing the values of X on the x axis and the probabilities $P(X)$ on the y axis.

Example 5–2

Represent graphically the probability distribution for the sample space for tossing three coins.

Number of heads X	0	1	2	3
Probability $P(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Solution

The values that X assumes are located on the x axis, and the values for $P(X)$ are located on the y axis. The graph is shown in Figure 5–1.

Note that for visual appearances, it is not necessary to start with 0 at the origin.

Examples 5–1 and 5–2 are illustrations of *theoretical* probability distributions. One did not need to actually perform the experiments to compute the probabilities. In contrast, to construct actual probability distributions, one must observe the variable over a period of time. They are empirical, as shown in Example 5–3.

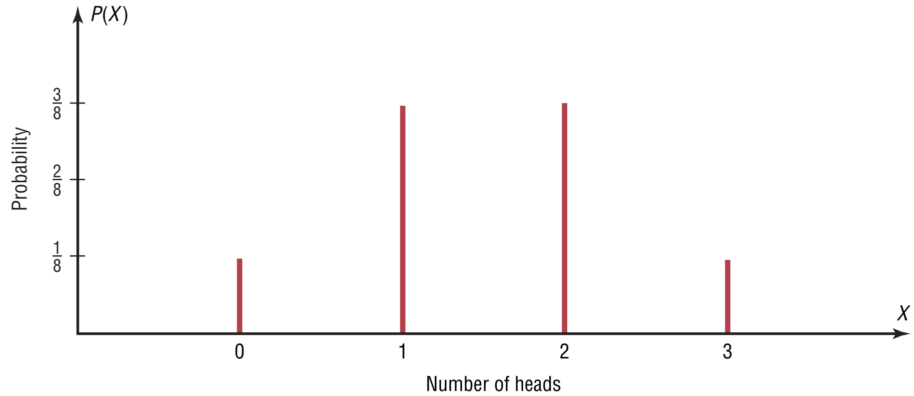


Figure 5–1
Probability Distribution for Example 5–2

Example 5–3

During the summer months, a rental agency keeps track of the number of chain saws it rents each day during a period of 90 days. The number of saws rented per day is represented by the variable *X*. The results are shown here. Compute the probability *P*(*X*) for each *X*, and construct a probability distribution and graph for the data.

X		**Number of days**
0		45
1		30
2		15
	Total	90

Solution

The probability *P*(*X*) can be computed for each *X* by dividing the number of days that *X* saws were rented by total days.

For 0 saws: $\frac{45}{90} = 0.50$

For 1 saw: $\frac{30}{90} = 0.33$

For 2 saws: $\frac{15}{90} = 0.17$

The distribution is shown here.

Number of saws rented *X*	0	1	2
Probability *P*(*X*)	0.50	0.33	0.17

The graph is shown in Figure 5–2.

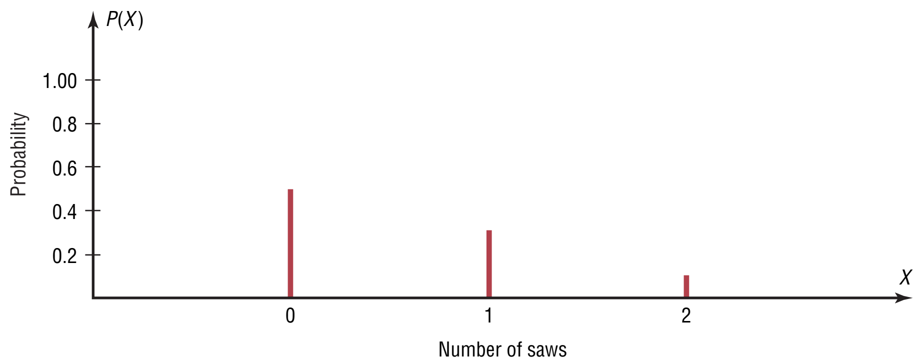


Figure 5–2
Probability Distribution for Example 5–3

Speaking of Statistics

Coins, Births, and Other Random (?) Events

Examples of random events such as tossing coins are used in almost all books on probability. But is flipping a coin really a random event?

Tossing coins dates back to ancient Roman times when the coins usually consisted of the Emperor's head on one side (i.e., heads) and another icon such as a ship on the other side (i.e., ships). Tossing coins was used in both fortune telling and ancient Roman games.

A Chinese form of divination called the *I-Ching* (pronounced E-Ching) is thought to be at least 4000 years old. It consists of 64 hexagrams made up of six horizontal lines. Each line is either broken or unbroken, representing the yin and the yang. These 64 hexagrams are supposed to represent all possible situations in life. To consult the I-Ching, a question is asked and then three coins are tossed six times. The way the coins fall, either heads up or heads down, determines whether the line is broken (yin) or unbroken (yang). Once the hexagon is determined, its meaning is consulted and interpreted to get the answer to the question. (*Note:* Another method used to determine the hexagon employs yarrow sticks.)

In the 16th century, a mathematician named Abraham DeMoivre used the outcomes of tossing coins to study what later became known as the normal distribution; however, his work at that time was not widely known.

Mathematicians usually consider the outcomes of a coin toss a random event. That is, each probability of getting a head is $\frac{1}{2}$, and the probability of getting a tail is $\frac{1}{2}$. Also, it is not possible to predict with 100% certainty which outcome will occur. But new studies question this theory. During World War II a South African mathematician named John Kerrich tossed a coin 10,000 times while he was interned in a German prison camp. Unfortunately, the results of his experiment were never recorded, so we don't know the number of heads that occurred.

Several studies have shown that when a coin-tossing device is used, the probability that a coin will land on the same side on which it is placed on the coin-tossing device is about 51%. It would take about 10,000 tosses to become aware of this bias. Furthermore, researchers showed that when a coin is spun on its edge, the coin would fall tails up about 80% of the time since there is more metal on the heads side of a coin. This makes the coin slightly heavier on the heads side than on the tails side.

Another assumption commonly made in probability theory is that the number of male births is equal to the number of female births and that the probability of a boy being born is $\frac{1}{2}$ and the probability of a girl being born is $\frac{1}{2}$. We know this is not exactly true.

In the later 1700s, a French mathematician named Pierre Simon Laplace attempted to prove that more males than females are born. He used records from 1745 to 1770 in Paris and showed that the percentage of females born was about 49%. Although these percentages vary somewhat from location to location, further surveys show they are generally true worldwide. Even though there are discrepancies, we generally consider the outcomes to be 50-50 since these discrepancies are relatively small.

Based on this article, would you consider the coin toss at the beginning of a football game fair?

Two Requirements for a Probability Distribution

1. The sum of the probabilities of all the events in the sample space must equal 1; that is, $\Sigma P(X) = 1$.
2. The probability of each event in the sample space must be between or equal to 0 and 1. That is, $0 \leq P(X) \leq 1$.

The first requirement states that the sum of the probabilities of all the events must be equal to 1. This sum cannot be less than 1 or greater than 1 since the sample space includes *all* possible outcomes of the probability experiment. The second requirement states that the probability of any individual event must be a value from 0 to 1. The reason (as stated in Chapter 4) is that the range of the probability of any individual value can be 0, 1, or any value between 0 and 1. A probability cannot be a negative number or greater than 1.

Example 5–4

Determine whether each distribution is a probability distribution.

a.

X	0	5	10	15	20
$P(X)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

b.

X	0	2	4	6
$P(X)$	-1.0	1.5	0.3	0.2

c.

X	1	2	3	4
$P(X)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{9}{16}$

d.

X	2	3	7
$P(X)$	0.5	0.3	0.4

Solution

a. Yes, it is a probability distribution.

b. No, it is not a probability distribution, since $P(X)$ cannot be 1.5 or -1.0.

c. Yes, it is a probability distribution.

d. No, it is not, since $\Sigma P(X) = 1.2$.

Many variables in business, education, engineering, and other areas can be analyzed by using probability distributions. Section 5–3 shows methods for finding the mean and standard deviation for a probability distribution.

Applying the Concepts 5–2

Dropping College Courses

Use the following table to answer the questions.

Reason for Dropping a College Course	Frequency	Percentage
Too difficult	45	
Illness	40	
Change in work schedule	20	
Change of major	14	
Family-related problems	9	
Money	7	
Miscellaneous	6	
No meaningful reason	3	

1. What is the variable under study? Is it a random variable?
2. How many people were in the study?
3. Complete the table.
4. From the information given, what is the probability that a student will drop a class because of illness? Money? Change of major?
5. Would you consider the information in the table to be a probability distribution?
6. Are the categories mutually exclusive?
7. Are the categories independent?
8. Are the categories exhaustive?
9. Are the two requirements for a discrete probability distribution met?

See page 283 for the answers.

Exercises 5–2

1. Define and give three examples of a random variable.

2. Explain the difference between a discrete and a continuous random variable.

3. Give three examples of a discrete random variable.

4. Give three examples of a continuous random variable.

5. What is a probability distribution? Give an example.

For Exercises 6 through 11, determine whether the distribution represents a probability distribution. If it does not, state why.

6.

X	1	6	11	16	21
$P(X)$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

7.

X	3	6	8	12
$P(X)$	0.3	0.5	0.7	−0.8

8.

X	3	6	8
$P(X)$	−0.3	0.6	0.7

9.

X	1	2	3	4	5
$P(X)$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

10.

X	20	30	40	50
$P(X)$	1.1	0.2	0.9	0.3

11.

X	5	10	15
$P(X)$	1.2	0.3	0.5

For Exercises 12 through 18, state whether the variable is discrete or continuous.

12. The speed of a jet airplane

13. The number of cheeseburgers a fast-food restaurant serves each day

14. The number of people who play the state lottery each day

15. The weight of a Siberian tiger

16. The time it takes to complete a marathon

17. The number of mathematics majors in your school

18. The blood pressures of all patients admitted to a hospital on a specific day

For Exercises 19 through 26, construct a probability distribution for the data and draw a graph for the distribution.

19. The probabilities that a patient will have 0, 1, 2, or 3 medical tests performed on entering a hospital are $\frac{6}{15}$, $\frac{5}{15}$, $\frac{3}{15}$, and $\frac{1}{15}$, respectively.

20. The probabilities of a return on an investment of \$1000, \$2000, and \$3000 are $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{4}$, respectively.

21. The probabilities of a machine manufacturing 0, 1, 2, 3, 4, or 5 defective parts in one day are 0.75, 0.17, 0.04, 0.025, 0.01, and 0.005, respectively.

22. The probabilities that a customer will purchase 0, 1, 2, or 3 books are 0.45, 0.30, 0.15, and 0.10, respectively.

23. A die is loaded in such a way that the probabilities of getting 1, 2, 3, 4, 5, and 6 are $\frac{1}{2}$, $\frac{1}{6}$, $\frac{1}{12}$, $\frac{1}{12}$, $\frac{1}{12}$, and $\frac{1}{12}$, respectively.

24. The probabilities that a customer selects 1, 2, 3, 4, and 5 items at a convenience store are 0.32, 0.12, 0.23, 0.18, and 0.15, respectively.

25. The probabilities that a surgeon operates on 3, 4, 5, 6, or 7 patients in any one day are 0.15, 0.20, 0.25, 0.20, and 0.20, respectively.

26. Three patients are given a headache relief tablet. The probabilities for 0, 1, 2, or 3 successes are 0.18, 0.52, 0.21, and 0.09, respectively.

27. A box contains two \$1 bills, three \$5 bills, one \$10 bill, and three \$20 bills. Construct a probability distribution for the data.

28. Construct a probability distribution for a family of three children. Let X represent the number of boys.

29. Construct a probability distribution for drawing a card from a deck of 40 cards consisting of 10 cards numbered 1, 10 cards numbered 2, 15 cards numbered 3, and 5 cards numbered 4.

30. Using the sample space for tossing two dice, construct a probability distribution for the sums 2 through 12.

Extending the Concepts

A probability distribution can be written in formula notation such as $P(X) = 1/X$, where $X = 2, 3, 6$. The distribution is shown as follows:

X	2	3	6
$P(X)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

For Exercises 31 through 36, write the distribution for the formula and determine whether it is a probability distribution.

31. $P(X) = X/6$ for $X = 1, 2, 3$

32. $P(X) = X$ for $X = 0.2, 0.3, 0.5$

33. $P(X) = X/6$ for $X = 3, 4, 7$

34. $P(X) = X + 0.1$ for $X = 0.1, 0.02, 0.04$

35. $P(X) = X/7$ for $X = 1, 2, 4$

36. $P(X) = X/(X + 2)$ for $X = 0, 1, 2$

5–3 Mean, Variance, Standard Deviation, and Expectation

Objective 2

Find the mean, variance, standard deviation, and expected value for a discrete random variable.

The mean, variance, and standard deviation for a probability distribution are computed differently from the mean, variance, and standard deviation for samples. This section explains how these measures—as well as a new measure called the *expectation*—are calculated for probability distributions.

Mean

In Chapter 3, the mean for a sample or population was computed by adding the values and dividing by the total number of values, as shown in the formulas

$$\overline{X} = \frac{\Sigma X}{n} \qquad \mu = \frac{\Sigma X}{N}$$

But how would one compute the mean of the number of spots that show on top when a die is rolled? One could try rolling the die, say, 10 times, recording the number of spots, and finding the mean; however, this answer would only approximate the true mean. What about 50 rolls or 100 rolls? Actually, the more times the die is rolled, the better the approximation. One might ask, then, How many times must the die be rolled to get the exact answer? *It must be rolled an infinite number of times.* Since this task is impossible, the previous formulas cannot be used because the denominators would be infinity. Hence, a new method of computing the mean is necessary. This method gives the exact theoretical value of the mean as if it were possible to roll the die an infinite number of times.

Historical Note

A professor, Augustin Louis Cauchy (1789–1857), wrote a book on probability. While he was teaching at the Military School of Paris, one of his students was Napoleon Bonaparte.

Before the formula is stated, an example will be used to explain the concept. Suppose two coins are tossed repeatedly, and the number of heads that occurred is recorded. What will be the mean of the number of heads? The sample space is

HH, HT, TH, TT

and each outcome has a probability of $\frac{1}{4}$. Now, in the long run, one would *expect* two heads (HH) to occur approximately $\frac{1}{4}$ of the time, one head to occur approximately $\frac{1}{2}$ of the time (HT or TH), and no heads (TT) to occur approximately $\frac{1}{4}$ of the time. Hence, on average, one would expect the number of heads to be

$$\tfrac{1}{4} \cdot 2 + \tfrac{1}{2} \cdot 1 + \tfrac{1}{4} \cdot 0 = 1$$

That is, if it were possible to toss the coins many times or an infinite number of times, the *average* of the number of heads would be 1.

Hence, to find the mean for a probability distribution, one must multiply each possible outcome by its corresponding probability and find the sum of the products.

Formula for the Mean of a Probability Distribution

The mean of a random variable with a discrete probability distribution is

$$\begin{aligned}\mu &= X_1 \cdot P(X_1) + X_2 \cdot P(X_2) + X_3 \cdot P(X_3) + \cdots + X_n \cdot P(X_n)\\ &= \Sigma X \cdot P(X)\end{aligned}$$

where $X_1, X_2, X_3, \ldots, X_n$ are the outcomes and $P(X_1), P(X_2), P(X_3), \ldots, P(X_n)$ are the corresponding probabilities.

Note: $\Sigma X \cdot P(X)$ means to sum the products.

Rounding Rule for the Mean, Variance, and Standard Deviation for a Probability Distribution The rounding rule for the mean, variance, and standard deviation for variables of a probability distribution is this: The mean, variance, and standard deviation should be rounded to one more decimal place than the outcome X. When fractions are used, they should be reduced to lowest terms.

Examples 5–5 through 5–8 illustrate the use of the formula.

Example 5–5 Find the mean of the number of spots that appear when a die is tossed.

Solution

In the toss of a die, the mean can be computed thus.

Outcome X	1	2	3	4	5	6
Probability $P(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\begin{aligned}\mu &= \Sigma X \cdot P(X) = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6}\\ &= \tfrac{21}{6} = 3\tfrac{1}{2} \text{ or } 3.5\end{aligned}$$

That is, when a die is tossed many times, the theoretical mean will be 3.5. Note that even though the die cannot show a 3.5, the theoretical average is 3.5.

The reason why this formula gives the theoretical mean is that in the long run, each outcome would occur approximately $\frac{1}{6}$ of the time. Hence, multiplying the outcome by its corresponding probability and finding the sum would yield the theoretical mean. In other words, outcome 1 would occur approximately $\frac{1}{6}$ of the time, outcome 2 would occur approximately $\frac{1}{6}$ of the time, etc.

Example 5–6

In a family with two children, find the mean of the number of children who will be girls.

Solution

The probability distribution is as follows:

Number of girls *X*	0	1	2
Probability *P*(*X*)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Hence, the mean is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1$$

Example 5–7

If three coins are tossed, find the mean of the number of heads that occur. (See the table preceding Example 5–1.)

Solution

The probability distribution is

Number of heads *X*	0	1	2	3
Probability *P*(*X*)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

The mean is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = \tfrac{12}{8} = 1\tfrac{1}{2} \text{ or } 1.5$$

The value 1.5 cannot occur as an outcome. Nevertheless, it is the long-run or theoretical average.

Example 5–8

The probability distribution shown represents the number of trips of five nights or more that American adults take per year. (That is, 6% do not take any trips lasting five nights or more, 70% take one trip lasting five nights or more per year, etc.) Find the mean.

Number of trips *X*	0	1	2	3	4
Probability *P*(*X*)	0.06	0.70	0.20	0.03	0.01

Solution

$$\begin{aligned}\mu &= \Sigma X \cdot P(X)\\ &= (0)(0.06) + (1)(0.70) + (2)(0.20) + (3)(0.03) + (4)(0.01)\\ &= 0 + 0.70 + 0.40 + 0.09 + 0.04\\ &= 1.23 \approx 1.2\end{aligned}$$

Hence, the mean of the number of trips lasting five nights or more per year taken by American adults is 1.2.

Historical Note

Fey Manufacturing Co., located in San Francisco, invented the first three-reel, automatic payout slot machine in 1895.

Variance and Standard Deviation

For a probability distribution, the mean of the random variable describes the measure of the so-called long-run or theoretical average, but it does not tell anything about the spread of the distribution. Recall from Chapter 3 that in order to measure this spread or variability, statisticians use the variance and standard deviation. These formulas were used:

$$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N} \qquad \text{or} \qquad \sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$

These formulas cannot be used for a random variable of a probability distribution since N is infinite, so the variance and standard deviation must be computed differently.

To find the variance for the random variable of a probability distribution, subtract the theoretical mean of the random variable from each outcome and square the difference. Then multiply each difference by its corresponding probability and add the products. The formula is

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

Finding the variance by using this formula is somewhat tedious. So for simplified computations, a shortcut formula can be used. This formula is algebraically equivalent to the longer one and is used in the examples that follow.

Formula for the Variance of a Probability Distribution

Find the variance of a probability distribution by multiplying the square of each outcome by its corresponding probability, summing those products, and subtracting the square of the mean. The formula for the variance of a probability distribution is

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

The standard deviation of a probability distribution is

$$\sigma = \sqrt{\sigma^2} \qquad \text{or} \qquad \sqrt{\Sigma[X^2 \cdot P(X)] - \mu^2}$$

Remember that the variance and standard deviation cannot be negative.

Example 5–9

Compute the variance and standard deviation for the probability distribution in Example 5–5.

Solution

Recall that the mean is $\mu = 3.5$, as computed in Example 5–5. Square each outcome and multiply by the corresponding probability, sum those products, and then subtract the square of the mean.

$$\sigma^2 = (1^2 \cdot \tfrac{1}{6} + 2^2 \cdot \tfrac{1}{6} + 3^2 \cdot \tfrac{1}{6} + 4^2 \cdot \tfrac{1}{6} + 5^2 \cdot \tfrac{1}{6} + 6^2 \cdot \tfrac{1}{6}) - (3.5)^2 = 2.9$$

To get the standard deviation, find the square root of the variance.

$$\sigma = \sqrt{2.9} = 1.7$$

Example 5–10

Five balls numbered 0, 2, 4, 6, and 8 are placed in a bag. After the balls are mixed, one is selected, its number is noted, and then it is replaced. If this experiment is repeated many times, find the variance and standard deviation of the numbers on the balls.

Historical Note

In 1657 a Dutch mathematician, Huygens, wrote a treatise on the Pascal-Fermat correspondence and introduced the idea of *mathematical expectation*.

Solution

Let X be the number on each ball. The probability distribution is

Number on ball X	0	2	4	6	8
Probability $P(X)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

The mean is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{5} + 2 \cdot \tfrac{1}{5} + 4 \cdot \tfrac{1}{5} + 6 \cdot \tfrac{1}{5} + 8 \cdot \tfrac{1}{5} = 4.0$$

The variance is

$$\begin{aligned}\sigma^2 &= \Sigma[X^2 \cdot P(X)] - \mu^2 \\ &= [0^2 \cdot (\tfrac{1}{5}) + 2^2 \cdot (\tfrac{1}{5}) + 4^2 \cdot (\tfrac{1}{5}) + 6^2 \cdot (\tfrac{1}{5}) + 8^2 \cdot (\tfrac{1}{5})] - 4^2 \\ &= [0 + \tfrac{4}{5} + \tfrac{16}{5} + \tfrac{36}{5} + \tfrac{64}{5}] - 16 \\ &= \tfrac{120}{5} - 16 \\ &= 24 - 16 = 8\end{aligned}$$

The standard deviation is $\sigma = \sqrt{8} = 2.8$.

The mean, variance, and standard deviation can also be found by using vertical columns, as shown [0.2 is used for $P(X)$ since $\frac{1}{5} = 0.2$].

X	$P(X)$	$X \cdot P(X)$	$X^2 \cdot P(X)$
0	0.2	0	0
2	0.2	0.4	0.8
4	0.2	0.8	3.2
6	0.2	1.2	7.2
8	0.2	1.6	12.8
		$\Sigma X \cdot P(X) = 4.0$	$\Sigma X^2 \cdot P(X) = 24.0$

Find the mean by summing the $X \cdot P(X)$ column and the variance by summing the $X^2 \cdot P(X)$ column and subtracting the square of the mean:

$$\sigma^2 = 24 - 4^2 = 8 \qquad \text{and} \qquad \sigma = \sqrt{8} = 2.8$$

Example 5–11

A talk radio station has four telephone lines. If the host is unable to talk (i.e., during a commercial) or is talking to a person, the other callers are placed on hold. When all lines are in use, others who are trying to call in get a busy signal. The probability that 0, 1, 2, 3, or 4 people will get through is shown in the distribution. Find the variance and standard deviation for the distribution.

X	0	1	2	3	4
$P(X)$	0.18	0.34	0.23	0.21	0.04

Should the station have considered getting more phone lines installed?

Solution

The mean is

$$\begin{aligned}\mu &= \Sigma X \cdot P(X)\\ &= 0 \cdot (0.18) + 1 \cdot (0.34) + 2 \cdot (0.23) + 3 \cdot (0.21) + 4 \cdot (0.04)\\ &= 1.6\end{aligned}$$

The variance is

$$\begin{aligned}\sigma^2 &= \Sigma[X^2 \cdot P(X)] - \mu^2\\ &= [0^2 \cdot (0.18) + 1^2 \cdot (0.34) + 2^2 \cdot (0.23) + 3^2 \cdot (0.21) + 4^2 \cdot (0.04)] - 1.6^2\\ &= [0 + 0.34 + 0.92 + 1.89 + 0.64] - 2.56\\ &= 3.79 - 2.56 = 1.23\\ &= 1.2 \text{ (rounded)}\end{aligned}$$

The standard deviation is $\sigma = \sqrt{\sigma^2}$, or $\sigma = \sqrt{1.2} = 1.1$.

No. The mean number of people calling at any one time is 1.6. Since the standard deviation is 1.1, most callers would be accommodated by having four phone lines because $\mu + 2\sigma$ would be $1.6 + 2(1.1) = 1.6 + 2.2 = 3.8$. Very few callers would get a busy signal since at least 75% of the callers would either get through or be put on hold. (See Chebyshev's theorem in Section 3–3.)

Expectation

Another concept related to the mean for a probability distribution is the concept of expected value or expectation. Expected value is used in various types of games of chance, in insurance, and in other areas, such as decision theory.

The **expected value** of a discrete random variable of a probability distribution is the theoretical average of the variable. The formula is

$$\mu = E(X) = \Sigma X \cdot P(X)$$

The symbol $E(X)$ is used for the expected value.

The formula for the expected value is the same as the formula for the theoretical mean. The expected value, then, is the theoretical mean of the probability distribution. That is, $E(X) = \mu$.

When expected value problems involve money, it is customary to round the answer to the nearest cent.

Example 5–12

One thousand tickets are sold at $1 each for a color television valued at $350. What is the expected value of the gain if a person purchases one ticket?

Solution

The problem can be set up as follows:

	Win	Lose
Gain X	$349	−$1
Probability $P(X)$	$\frac{1}{1000}$	$\frac{999}{1000}$

Two things should be noted. First, for a win, the net gain is \$349, since the person does not get the cost of the ticket (\$1) back. Second, for a loss, the gain is represented by a negative number, in this case −\$1. The solution, then, is

$$E(X) = \$349 \cdot \frac{1}{1000} + (-\$1) \cdot \frac{999}{1000} = -\$0.65$$

Expected value problems of this type can also be solved by finding the overall gain (i.e., the value of the prize won or the amount of money won, not considering the cost of the ticket for the prize or the cost to play the game) and subtracting the cost of the tickets or the cost to play the game, as shown:

$$E(X) = \$350 \cdot \frac{1}{1000} - \$1 = -\$0.65$$

Here, the overall gain (\$350) must be used.

Note that the expectation is −\$0.65. This does not mean that a person loses \$0.65, since the person can only win a television set valued at \$350 or lose \$1 on the ticket. What this expectation means is that the average of the losses is \$0.65 for each of the 1000 ticket holders. Here is another way of looking at this situation: If a person purchased one ticket each week over a long time, the average loss would be \$0.65 per ticket, since theoretically, on average, that person would win the set once for each 1000 tickets purchased.

Example 5–13

One thousand tickets are sold at \$1 each for four prizes of \$100, \$50, \$25, and \$10. What is the expected value if a person purchases two tickets?

Gain *X*	\$98	\$48	\$23	\$8	−\$2
Probability *P*(*X*)	$\frac{2}{1000}$	$\frac{2}{1000}$	$\frac{2}{1000}$	$\frac{2}{1000}$	$\frac{992}{1000}$

Solution

$$E(X) = \$98 \cdot \frac{2}{1000} + \$48 \cdot \frac{2}{1000} + \$23 \cdot \frac{2}{1000} + \$8 \cdot \frac{2}{1000} + (-\$2) \cdot \frac{992}{1000}$$
$$= -\$1.63$$

An alternate solution is

$$E(X) = \$100 \cdot \frac{2}{1000} + \$50 \cdot \frac{2}{1000} + \$25 \cdot \frac{2}{1000} + \$10 \cdot \frac{2}{1000} - \$2$$
$$= -\$1.63$$

Example 5–14

A financial adviser suggests that his client select one of two types of bonds in which to invest \$5000. Bond *X* pays a return of 4% and has a default rate of 2%. Bond *Y* has a $2\frac{1}{2}$% return and a default rate of 1%. Find the expected rate of return and decide which bond would be a better investment. When the bond defaults, the investor loses all the investment.

Solution

The return on bond X is $\$5000 \cdot 4\% = \$200$. The expected return then is

$$E(X) = \$200(0.98) - \$5000(0.02) = \$96$$

The return on bond Y is $\$5000 \cdot 2\frac{1}{2}\% = \$125$. The expected return then is

$$E(X) = \$125(0.99) - \$5000(0.01) = \$73.75$$

Hence, bond X would be a better investment since the expected return is higher.

In gambling games, if the expected value of the game is zero, the game is said to be fair. If the expected value of a game is positive, then the game is in favor of the player. That is, the player has a better-than-even chance of winning. If the expected value of the game is negative, then the game is said to be in favor of the house. That is, in the long run, the players will lose money.

In his book *Probabilities in Everyday Life* (Ivy Books, 1986), author John D. McGervy gives the expectations for various casino games. For keno, the house wins \$0.27 on every \$1.00 bet. For Chuck-a-Luck, the house wins about \$0.52 on every \$1.00 bet. For roulette, the house wins about \$0.90 on every \$1.00 bet. For craps, the house wins about \$0.88 on every \$1.00 bet. The bottom line here is that if you gamble long enough, sooner or later you will end up losing money.

Applying the Concepts 5–3

Expected Value

On March 28, 1979, the nuclear generating facility at Three Mile Island, Pennsylvania, began discharging radiation into the atmosphere. People exposed to even low levels of radiation can experience health problems ranging from very mild to severe, even causing death. A local newspaper reported that 11 babies were born with kidney problems in the three-county area surrounding the Three Mile Island nuclear power plant. The expected value for that problem in infants in that area was 3. Answer the following questions.

1. What does *expected value* mean?
2. Would you expect the exact value of 3 all the time?
3. If a news reporter stated that the number of cases of kidney problems in newborns was nearly four times as much as was usually expected, do you think pregnant mothers living in that area would be overly concerned?
4. Is it unlikely that 11 occurred by chance?
5. Are there any other statistics that could better inform the public?
6. Assume that 3 out of 2500 babies were born with kidney problems in that three-county area the year before the accident. Also assume that 11 out of 2500 babies were born with kidney problems in that three-county area the year after the accident. What is the real percent of increase in that abnormality?
7. Do you think that pregnant mothers living in that area should be overly concerned after looking at the results in terms of rates?

See page 283 for the answers.

Exercises 5–3

1. From past experience, a company has found that in cartons of transistors, 92% contain no defective transistors, 3% contain one defective transistor, 3% contain two defective transistors, and 2% contain three defective transistors. Find the mean, variance, and standard deviation for the defective transistors.

About how many extra transistors per day would the company need to replace the defective ones if it used 10 cartons per day?

2. The number of suits sold per day at a retail store is shown in the table, with the corresponding probabilities. Find the mean, variance, and standard deviation of the distribution.

Number of suits sold ***X***	19	20	21	22	23
Probability ***P*(*X*)**	0.2	0.2	0.3	0.2	0.1

If the manager of the retail store wants to be sure that he has enough suits for the next 5 days, how many should the manager purchase?

3. A bank vice president feels that each savings account customer has, on average, three credit cards. The following distribution represents the number of credit cards people own. Find the mean, variance, and standard deviation. Is the vice president correct?

Number of cards ***X***	0	1	2	3	4
Probability ***P*(*X*)**	0.18	0.44	0.27	0.08	0.03

4. The number of refrigerators sold per day at a local appliance store is shown in the table, along with the corresponding probabilities. Find the mean, variance, and standard deviation.

Number of refrigerators sold ***X***	0	1	2	3	4
Probability ***P*(*X*)**	0.1	0.2	0.3	0.2	0.2

5. A public speaker computes the probabilities for the number of speeches she gives each week. Compute the mean, variance, and standard deviation of the distribution shown.

Number of speeches ***X***	0	1	2	3	4	5
Probability ***P*(*X*)**	0.06	0.42	0.22	0.12	0.15	0.03

If she receives $100 per speech, about how much will she earn per week?

6. A recent survey by an insurance company showed the following probabilities for the number of bedrooms in each insured home. Find the mean, variance, and standard deviation for the distribution.

Number of bedrooms ***X***	2	3	4	5
Probability ***P*(*X*)**	0.3	0.4	0.2	0.1

7. A concerned parents group determined the number of commercials shown in each of five children's programs over a period of time. Find the mean, variance, and standard deviation for the distribution shown.

Number of commercials ***X***	5	6	7	8	9
Probability ***P*(*X*)**	0.2	0.25	0.38	0.10	0.07

8. A study conducted by a TV station showed the number of televisions per household and the corresponding probabilities for each. Find the mean, variance, and standard deviation.

Number of televisions ***X***	1	2	3	4
Probability ***P*(*X*)**	0.32	0.51	0.12	0.05

If you were taking a survey on the programs that were watched on television, how many program diaries would you send to each household in the survey?

9. The following distribution shows the number of students enrolled in CPR classes offered by the local fire department. Find the mean, variance, and standard deviation for the distribution.

Number of students ***X***	12	13	14	15	16
Probability ***P*(*X*)**	0.15	0.20	0.38	0.18	0.09

10. A pizza shop owner determines the number of pizzas that are delivered each day. Find the mean, variance, and standard deviation for the distribution shown. If the manager stated that 45 pizzas were delivered on one day, do you think that this is a believable claim?

Number of deliveries ***X***	35	36	37	38	39
Probability ***P*(*X*)**	0.1	0.2	0.3	0.3	0.1

11. An insurance company insures a person's antique coin collection worth $20,000 for an annual premium of $300. If the company figures that the probability of the collection being stolen is 0.002, what will be the company's expected profit?

12. A landscape contractor bids on jobs where he can make $3000 profit. The probabilities of getting one, two, three, or four jobs per month are shown.

Number of jobs	1	2	3	4
Probability	0.2	0.3	0.4	0.1

Find the contractor's expected profit per month.

13. If a person rolls doubles when he tosses two dice, he wins \$5. For the game to be fair, how much should the person pay to play the game?

14. If a player rolls two dice and gets a sum of 2 or 12, she wins \$20. If the person gets a 7, she wins \$5. The cost to play the game is \$3. Find the expectation of the game.

15. A lottery offers one \$1000 prize, one \$500 prize, and five \$100 prizes. One thousand tickets are sold at \$3 each. Find the expectation if a person buys one ticket.

16. In Exercise 15, find the expectation if a person buys two tickets. Assume that the player's ticket is replaced after each draw and that the same ticket can win more than one prize.

17. For a daily lottery, a person selects a three-digit number. If the person plays for \$1, she can win \$500. Find the expectation. In the same daily lottery, if a person boxes a number, she will win \$80. Find the expectation if the number 123 is played for \$1 and boxed. (When a number is "boxed," it can win when the digits occur in any order.)

18. A 35-year-old woman purchases a \$100,000 term life insurance policy for an annual payment of \$360. Based on a period life table for the U.S. government, the probability that she will survive the year is 0.999057. Find the expected value of the policy for the insurance company.

19. A person decides to invest \$50,000 in a gas well. Based on history, the probabilities of the outcomes are as follows.

Outcome		***P*(*X*)**
\$80,000	(Highly successful)	0.2
\$40,000	(Moderately successful)	0.7
−\$50,000	(Dry well)	0.1

Find the expected value of the investment. Would you consider this a good investment?

Extending the Concepts

20. Construct a probability distribution for the sum shown on the faces when two dice are rolled. Find the mean, variance, and standard deviation of the distribution.

21. When one die is rolled, the expected value of the number of spots is 3.5. In Exercise 20, the mean number of spots was found for rolling two dice. What is the mean number of spots if three dice are rolled?

22. The formula for finding the variance for a probability distribution is

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

Verify algebraically that this formula gives the same result as the shortcut formula shown in this section.

23. Roll a die 100 times. Compute the mean and standard deviation. How does the result compare with the theoretical results of Example 5–5?

24. Roll two dice 100 times and find the mean, variance, and standard deviation of the sum of the spots. Compare the result with the theoretical results obtained in Exercise 20.

25. Conduct a survey of the number of extracurricular activities your classmates are enrolled in. Construct a probability distribution and find the mean, variance, and standard deviation.

26. In a recent promotional campaign, a company offered these prizes and the corresponding probabilities. Find the expected value of winning. The tickets are free.

Number of prizes	**Amount**	**Probability**
1	\$100,000	$\frac{1}{1,000,000}$
2	10,000	$\frac{1}{50,000}$
5	1,000	$\frac{1}{10,000}$
10	100	$\frac{1}{1000}$

If the winner has to mail in the winning ticket to claim the prize, what will be the expectation if the cost of the stamp is considered? Use the current cost of a stamp for a first-class letter.

Speaking of Statistics

This study shows that a part of the brain reacts to the impact of losing, and it might explain why people tend to increase their bets after losing when gambling. Explain how this type of split decision making may influence fighter pilots, firefighters, or police officers, as the article states.

THE GAMBLER'S FALLACY

WHY WE EXPECT TO STRIKE IT RICH AFTER A LOSING STREAK

A GAMBLER USUALLY WAGERS more after taking a loss, in the misguided belief that a run of bad luck increases the probability of a win. We tend to cling to the misconception that past events can skew future odds. "On some level, you're thinking, 'If I just lost, it's going to even out.' The extent to which you're disturbed by a loss seems to go along with risky behavior," says University of Michigan psychologist William Gehring, Ph.D., co-author of a new study linking dicey decision-making to neurological activity originating in the medial frontal cortex, long thought to be an area of the brain used in error detection.

Because people are so driven to up the ante after a loss, Gehring believes that the medial frontal cortex unconsciously influences future decisions based on the impact of the loss, in addition to registering the loss itself.

Gehring drew this conclusion by asking 12 subjects fitted with electrode caps to choose either the number 5 or 25, with the larger number representing the riskier bet. On any given round, both numbers could amount to a loss, both could amount to a gain or the results could split, one number signifying a loss, the other a gain.

The medial frontal cortex responded to the outcome of a gamble within a quarter of a second, registering sharp electrical impulses only after a loss. Gehring points out that if the medial frontal cortex simply detected errors it would have reacted after participants chose the lesser of two possible gains. In other words, choosing "5" during a round in which both numbers paid off and betting on "25" would have yielded a larger profit.

After the study appeared in *Science*, Gehring received several e-mails from stock traders likening the "gambler's fallacy" to impulsive trading decisions made directly after off-loading a losing security. Researchers speculate that such risky, split-second decision-making could extend to fighter pilots, firemen and policemen—professions in which rapid-fire decisions are crucial and frequent.

—Dan Schulman

Source: Psychology Today, August 2002, p. 22. Used with permission.

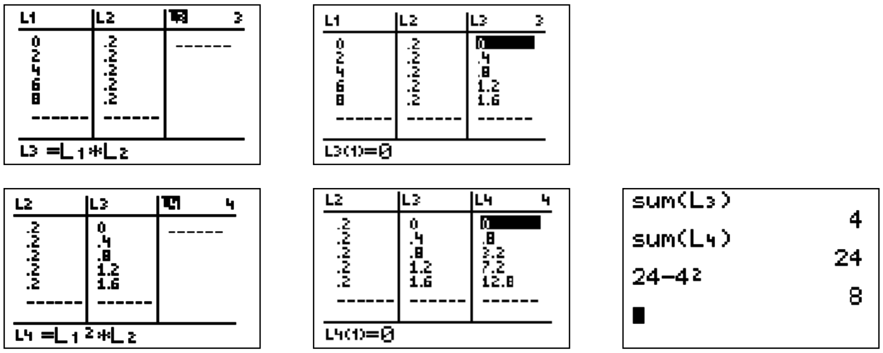
Technology *Step by Step*

TI-83 Plus or TI-84 Plus Step by Step

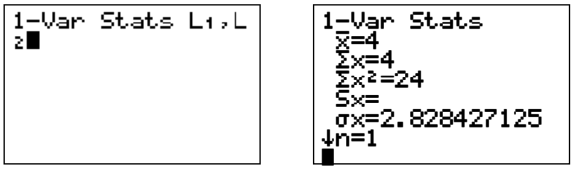
To calculate the mean and variance for a discrete random variable by using the formulas:

1. Enter the x values into L_1 and the probabilities into L_2.
2. Move the cursor to the top of the L_3 column so that L_3 is highlighted.
3. Type $\mathbf{L_1}$ multiplied by $\mathbf{L_2}$, then press **ENTER.**
4. Move the cursor to the top of the L_4 column so that L_4 is highlighted.
5. Type $\mathbf{L_1}$ followed by the $\boldsymbol{x^2}$ key multiplied by $\mathbf{L_2}$, then press **ENTER.**
6. Type **2nd QUIT** to return to the home screen.
7. Type **2nd LIST,** move the cursor to MATH, type **5** for sum(, then type $\mathbf{L_3}$) then press **ENTER.**
8. Type **2nd ENTER,** move the cursor to L_3, type L_4, then press **ENTER.**

Using the data from Example 5–10 gives the following:

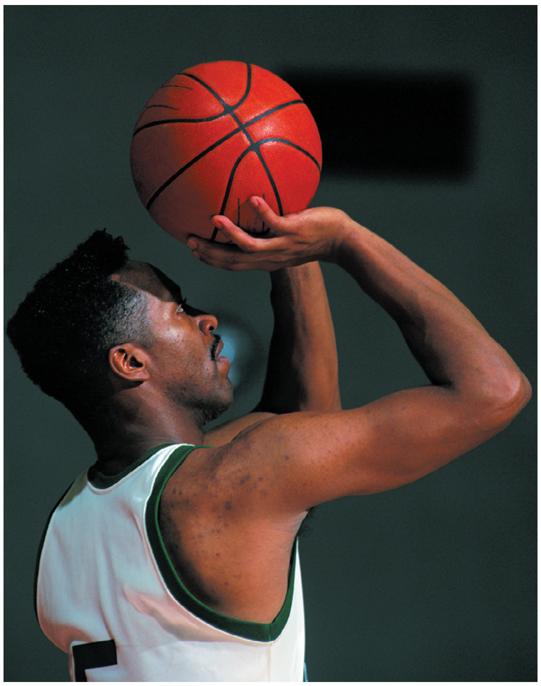


To calculate the mean and standard deviation for a discrete random variable without using the formulas, modify the procedure to calculate the mean and standard deviation from grouped data (Chapter 3) by entering the x values into L_1 and the probabilities into L_2.



5–4 The Binomial Distribution

Many types of probability problems have only two outcomes or can be reduced to two outcomes. For example, when a coin is tossed, it can land heads or tails. When a baby is born, it will be either male or female. In a basketball game, a team either wins or loses. A true/false item can be answered in only two ways, true or false. Other situations can be



Objective 3

Find the exact probability for X successes in n trials of a binomial experiment.

Historical Note

In 1653, Blaise Pascal created a triangle of numbers called *Pascal's triangle* that can be used in the binomial distribution.

reduced to two outcomes. For example, a medical treatment can be classified as effective or ineffective, depending on the results. A person can be classified as having normal or abnormal blood pressure, depending on the measure of the blood pressure gauge. A multiple-choice question, even though there are four or five answer choices, can be classified as correct or incorrect. Situations like these are called *binomial experiments.*

A **binomial experiment** is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. These outcomes can be considered as either success or failure.
3. The outcomes of each trial must be independent of each other.
4. The probability of a success must remain the same for each trial.

A binomial experiment and its results give rise to a special probability distribution called the *binomial distribution.*

The outcomes of a binomial experiment and the corresponding probabilities of these outcomes are called a **binomial distribution.**

In binomial experiments, the outcomes are usually classified as successes or failures. For example, the correct answer to a multiple-choice item can be classified as a success, but any of the other choices would be incorrect and hence classified as a failure. The notation that is commonly used for binomial experiments and the binomial distribution is defined now.

Notation for the Binomial Distribution

$P(S)$	The symbol for the probability of success
$P(F)$	The symbol for the probability of failure
p	The numerical probability of a success
q	The numerical probability of a failure

$$P(S) = p \qquad \text{and} \qquad P(F) = 1 - p = q$$

n	The number of trials
X	The number of successes in n trials

Note that $0 \leq X \leq n$ and $X = 0, 1, 2, 3, \ldots, n$

The probability of a success in a binomial experiment can be computed with this formula.

Binomial Probability Formula

In a binomial experiment, the probability of exactly X successes in n trials is

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

An explanation of why the formula works is given following Example 5–15.

Example 5–15

A coin is tossed 3 times. Find the probability of getting exactly two heads.

Solution

This problem can be solved by looking at the sample space. There are three ways to get two heads.

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

The answer is $\frac{3}{8}$, or 0.375.

Looking at the problem in Example 5–15 from the standpoint of a binomial experiment, one can show that it meets the four requirements.

1. There are a fixed number of trials (three).
2. There are only two outcomes for each trial, heads or tails.
3. The outcomes are independent of one another (the outcome of one toss in no way affects the outcome of another toss).
4. The probability of a success (heads) is $\frac{1}{2}$ in each case.

In this case, $n = 3$, $X = 2$, $p = \frac{1}{2}$, and $q = \frac{1}{2}$. Hence, substituting in the formula gives

$$P(\text{2 heads}) = \frac{3!}{(3-2)!2!} \cdot \left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8} = 0.375$$

which is the same answer obtained by using the sample space.

The same example can be used to explain the formula. First, note that there are three ways to get exactly two heads and one tail from a possible eight ways. They are HHT, HTH, and THH. In this case, then, the number of ways of obtaining two heads from three coin tosses is ${}_3C_2$, or 3, as shown in Chapter 4. In general, the number of ways to get X successes from n trials without regard to order is

$${}_nC_X = \frac{n!}{(n-X)!X!}$$

This is the first part of the binomial formula. (Some calculators can be used for this.)

Next, each success has a probability of $\frac{1}{2}$ and can occur twice. Likewise, each failure has a probability of $\frac{1}{2}$ and can occur once, giving the $(\frac{1}{2})^2(\frac{1}{2})^1$ part of the formula. To generalize, then, each success has a probability of p and can occur X times, and each failure has a probability of q and can occur $n - X$ times. Putting it all together yields the binomial probability formula.

Example 5–16

A survey found that one out of five Americans say he or she has visited a doctor in any given month. If 10 people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

Source: Reader's Digest.

Solution

In this case, $n = 10$, $X = 3$, $p = \frac{1}{5}$, and $q = \frac{4}{5}$. Hence,

$$P(3) = \frac{10!}{(10-3)!3!}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7 = 0.201$$

Example 5–17

A survey from Teenage Research Unlimited (Northbrook, Illinois) found that 30% of teenage consumers receive their spending money from part-time jobs. If 5 teenagers are selected at random, find the probability that at least 3 of them will have part-time jobs.

Solution

To find the probability that at least 3 have part-time jobs, it is necessary to find the individual probabilities for 3, or 4, or 5, and then add them to get the total probability.

$$P(3) = \frac{5!}{(5-3)!3!}(0.3)^3(0.7)^2 = 0.132$$

$$P(4) = \frac{5!}{(5-4)!4!}(0.3)^4(0.7)^1 = 0.028$$

$$P(5) = \frac{5!}{(5-5)!5!}(0.3)^5(0.7)^0 = 0.002$$

Hence,

$$P(\text{at least three teenagers have part-time jobs}) = 0.132 + 0.028 + 0.002 = 0.162$$

Computing probabilities by using the binomial probability formula can be quite tedious at times, so tables have been developed for selected values of n and p. Table B in Appendix C gives the probabilities for individual events. Example 5–18 shows how to use Table B to compute probabilities for binomial experiments.

Example 5–18

Solve the problem in Example 5–15 by using Table B.

Solution

Since $n = 3$, $X = 2$, and $p = 0.5$, the value 0.375 is found as shown in Figure 5–3.

Figure 5–3
Using Table B for Example 5–18

		p										
n	X	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2	0											
	1											
	2											
3	0						0.125					
	1						0.375					
	2						0.375					
	3						0.125					

$p = 0.5$

$n = 3$

$X = 2$

Example 5–19

Public Opinion reported that 5% of Americans are afraid of being alone in a house at night. If a random sample of 20 Americans is selected, find these probabilities by using the binomial table.

a. There are exactly 5 people in the sample who are afraid of being alone at night.

b. There are at most 3 people in the sample who are afraid of being alone at night.

c. There are at least 3 people in the sample who are afraid of being alone at night.

Source: 100% American by Daniel Evan Weiss.

Solution

a. $n = 20$, $p = 0.05$, and $X = 5$. From the table, one gets 0.002.

b. $n = 20$ and $p = 0.05$. "At most 3 people" means 0, or 1, or 2, or 3.

Hence, the solution is

$$\begin{aligned} P(0) + P(1) + P(2) + P(3) &= 0.358 + 0.377 + 0.189 + 0.060 \\ &= 0.984 \end{aligned}$$

c. $n = 20$ and $p = 0.05$. "At least 3 people" means 3, 4, 5, . . . , 20. This problem can best be solved by finding $P(0) + P(1) + P(2)$ and subtracting from 1.

$$P(0) + P(1) + P(2) = 0.358 + 0.377 + 0.189 = 0.924$$
$$1 - 0.924 = 0.076$$

Example 5–20

A report from the Secretary of Health and Human Services stated that 70% of single-vehicle traffic fatalities that occur at night on weekends involve an intoxicated driver. If a sample of 15 single-vehicle traffic fatalities that occur at night on a weekend is selected, find the probability that exactly 12 involve a driver who is intoxicated.

Source: 100% American by Daniel Evan Weiss.

Solution

Now, $n = 15$, $p = 0.70$, and $X = 12$. From Table B, $P(12) = 0.170$. Hence, the probability is 0.17.

Remember that in the use of the binomial distribution, the outcomes must be independent. For example, in the selection of components from a batch to be tested, each component must be replaced before the next one is selected. Otherwise, the outcomes are not independent. However, a dilemma arises because there is a chance that the same component could be selected again. This situation can be avoided by not replacing the component and using a distribution called the hypergeometric distribution to calculate the probabilities. The hypergeometric distribution is presented later in this chapter. Note that when the population is large and the sample is small, the binomial probabilities can be shown to be nearly the same as the corresponding hypergeometric probabilities.

Objective 4

Find the mean, variance, and standard deviation for the variable of a binomial distribution.

Mean, Variance, and Standard Deviation for the Binomial Distribution

The mean, variance, and standard deviation of a variable that has the *binomial distribution* can be found by using the following formulas.

Mean $\mu = n \cdot p$ Variance $\sigma^2 = n \cdot p \cdot q$ Standard deviation $\sigma = \sqrt{n \cdot p \cdot q}$

These formulas are algebraically equivalent to the formulas for the mean, variance, and standard deviation of the variables for probability distributions, but because they are for variables of the binomial distribution, they have been simplified by using algebra. The algebraic derivation is omitted here, but their equivalence is shown in Example 5–21.

Example 5–21

A coin is tossed 4 times. Find the mean, variance, and standard deviation of the number of heads that will be obtained.

Solution

With the formulas for the binomial distribution and $n = 4$, $p = \frac{1}{2}$, and $q = \frac{1}{2}$, the results are

$$\mu = n \cdot p = 4 \cdot \tfrac{1}{2} = 2$$
$$\sigma^2 = n \cdot p \cdot q = 4 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 1$$
$$\sigma = \sqrt{1} = 1$$

From Example 5–21, when four coins are tossed many, many times, the average of the number of heads that appear is 2, and the standard deviation of the number of heads is 1. Note that these are theoretical values.

As stated previously, this problem can be solved by using the formulas for expected value. The distribution is shown.

No. of heads X	0	1	2	3	4
Probability $P(X)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$\mu = E(X) = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{16} + 1 \cdot \tfrac{4}{16} + 2 \cdot \tfrac{6}{16} + 3 \cdot \tfrac{4}{16} + 4 \cdot \tfrac{1}{16} = \tfrac{32}{16} = 2$$
$$\sigma^2 = \Sigma X^2 \cdot P(X) - \mu^2$$
$$= 0^2 \cdot \tfrac{1}{16} + 1^2 \cdot \tfrac{4}{16} + 2^2 \cdot \tfrac{6}{16} + 3^2 \cdot \tfrac{4}{16} + 4^2 \cdot \tfrac{1}{16} - 2^2 = \tfrac{80}{16} - 4 = 1$$
$$\sigma = \sqrt{1} = 1$$

Hence, the simplified binomial formulas give the same results.

Example 5–22

A die is rolled 480 times. Find the mean, variance, and standard deviation of the number of 2s that will be rolled.

Solution

This is a binomial situation, where getting a 2 is a success and not getting a 2 is a failure; hence, $n = 480$, $p = \frac{1}{6}$, and $q = \frac{5}{6}$.

$$\mu = n \cdot p = 480 \cdot \tfrac{1}{6} = 80$$
$$\sigma^2 = n \cdot p \cdot q = 480 \cdot (\tfrac{1}{6})(\tfrac{5}{6}) = 66.7$$
$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{66.7} = 8.2$$

On average, there will be eighty 2s. The standard deviation is 8.2.

Example 5–23

The *Statistical Bulletin* published by Metropolitan Life Insurance Co. reported that 2% of all American births result in twins. If a random sample of 8000 births is taken, find the mean, variance, and standard deviation of the number of births that would result in twins.

Source: 100% American by Daniel Evan Weiss.

Solution

This is a binomial situation, since a birth can result in either twins or not twins (i.e., two outcomes).

$$\mu = n \cdot p = (8000)(0.02) = 160$$
$$\sigma^2 = n \cdot p \cdot q = (8000)(0.02)(0.98) = 156.8$$
$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{156.8} = 12.5$$

For the sample, the average number of births that would result in twins is 160, the variance is 156.8, or 157, and the standard deviation is 12.5, or 13 if rounded.

Applying the Concepts 5–4

Unsanitary Restaurants

Health officials routinely check sanitary conditions of restaurants. Assume you visit a popular tourist spot and read in the newspaper that in 3 out of every 7 restaurants checked, there were unsatisfactory health conditions found. Assuming you are planning to eat out 10 times while you are there on vacation, answer the following questions.

1. How likely is it that you will eat at three restaurants with unsanitary conditions?
2. How likely is it that you will eat at four or five restaurants with unsanitary conditions?
3. Explain how you would compute the probability of eating in at least one restaurant with unsanitary conditions. Could you use the complement to solve this problem?
4. What is the most likely number to occur in this experiment?
5. How variable will the data be around the most likely number?
6. Is this a binomial distribution?
7. If it is a binomial distribution, does that mean that the likelihood of a success is always 50% since there are only two possible outcomes?

Check your answers by using the following computer-generated table.

Mean = 4.3 **Std. dev. = 1.56557**

X	*P*(*X*)	Cum. Prob.
0	0.00362	0.00362
1	0.02731	0.03093
2	0.09272	0.12365
3	0.18651	0.31016
4	0.24623	0.55639
5	0.22291	0.77930
6	0.14013	0.91943
7	0.06041	0.97983
8	0.01709	0.99692
9	0.00286	0.99979
10	0.00022	1.00000

See page 283 for the answers.

Exercises 5–4

1. Which of the following are binomial experiments or can be reduced to binomial experiments?
   a. Surveying 100 people to determine if they like Sudsy Soap
   b. Tossing a coin 100 times to see how many heads occur
   c. Drawing a card with replacement from a deck and getting a heart
   d. Asking 1000 people which brand of cigarettes they smoke
   e. Testing four different brands of aspirin to see which brands are effective
   f. Testing one brand of aspirin by using 10 people to determine whether it is effective
   g. Asking 100 people if they smoke
   h. Checking 1000 applicants to see whether they were admitted to White Oak College
   i. Surveying 300 prisoners to see how many different crimes they were convicted of
   j. Surveying 300 prisoners to see whether this is their first offense

2. **(ans)** Compute the probability of X successes, using Table B in Appendix C.
   a. $n = 2, p = 0.30, X = 1$
   b. $n = 4, p = 0.60, X = 3$
   c. $n = 5, p = 0.10, X = 0$
   d. $n = 10, p = 0.40, X = 4$
   e. $n = 12, p = 0.90, X = 2$
   f. $n = 15, p = 0.80, X = 12$
   g. $n = 17, p = 0.05, X = 0$
   h. $n = 20, p = 0.50, X = 10$
   i. $n = 16, p = 0.20, X = 3$

3. Compute the probability of X successes, using the binomial formula.
   a. $n = 6, X = 3, p = 0.03$
   b. $n = 4, X = 2, p = 0.18$
   c. $n = 5, X = 3, p = 0.63$
   d. $n = 9, X = 0, p = 0.42$
   e. $n = 10, X = 5, p = 0.37$

For Exercises 4 through 13, assume all variables are binomial. (*Note:* If values are not found in Table B of Appendix C, use the binomial formula.)

4. A burglar alarm system has six fail-safe components. The probability of each failing is 0.05. Find these probabilities.
   a. Exactly three will fail.
   b. Fewer than two will fail.
   c. None will fail.
   d. Compare the answers for parts *a*, *b*, and *c*, and explain why the results are reasonable.

5. A student takes a 20-question, true/false exam and guesses on each question. Find the probability of passing if the lowest passing grade is 15 correct out of 20. Would you consider this event likely to occur? Explain your answer.

6. A student takes a 20-question, multiple-choice exam with five choices for each question and guesses on each question. Find the probability of guessing at least 15 out of 20 correctly. Would you consider this event likely or unlikely to occur? Explain your answer.

7. In a survey, 30% of the people interviewed said that they bought most of their books during the last 3 months of the year (October, November, December). If nine people are selected at random, find the probability that exactly three of these people bought most of their books during October, November, and December.
   Source: USA Snapshot, *USA TODAY.*

8. In a Gallup Survey, 90% of the people interviewed were unaware that maintaining a healthy weight could reduce the risk of stroke. If 15 people are selected at random, find the probability that at least 9 are unaware that maintaining a proper weight could reduce the risk of stroke.
   Source: USA Snapshot, *USA TODAY.*

9. In a survey, three of four students said the courts show “too much concern” for criminals. Find the probability that at most three out of seven randomly selected students will agree with this statement.
   Source: *Harper’s Index.*

10. It was found that 60% of American victims of health care fraud are senior citizens. If 10 victims are randomly selected, find the probability that exactly 3 are senior citizens.
    Source: *100% American* by Daniel Evan Weiss.

11. R. H. Bruskin Associates Market Research found that 40% of Americans do not think that having a college education is important to succeed in the business world. If a random sample of five Americans is selected, find these probabilities.
    a. Exactly two people will agree with that statement.
    b. At most three people will agree with that statement.
    c. At least two people will agree with that statement.
    d. Fewer than three people will agree with that statement.
    Source: *100% American* by Daniel Evans Weiss.

12. Find these probabilities for a sample of 20 teenagers if 70% of them had compact disk players by the age of 16.

a. At least 14 had CD players

b. Exactly 9 had CD players

c. More than 17 had CD players

d. Which event, *a*, *b*, or *c*, is most likely to occur? Explain why.

13. If 80% of the people in a community have Internet access from their homes, find these probabilities for a sample of 10 people.

a. At most 6 have Internet access.

b. Exactly 6 have Internet access.

c. At least 6 have Internet access.

d. Which event, *a*, *b*, or *c*, is most likely to occur? Explain why.

14. **(ans)** Find the mean, variance, and standard deviation for each of the values of n and p when the conditions for the binomial distribution are met.

a. $n = 100, p = 0.75$

b. $n = 300, p = 0.3$

c. $n = 20, p = 0.5$

d. $n = 10, p = 0.8$

e. $n = 1000, p = 0.1$

f. $n = 500, p = 0.25$

g. $n = 50, p = \frac{2}{5}$

h. $n = 36, p = \frac{1}{6}$

15. A study found that 1% of Social Security recipients are too young to vote. If 800 Social Security recipients are randomly selected, find the mean, variance, and standard deviation of the number of recipients who are too young to vote.

Source: *Harper's Index.*

16. Find the mean, variance, and standard deviation for the number of heads when 20 coins are tossed.

17. If 3% of calculators are defective, find the mean, variance, and standard deviation of a lot of 300 calculators.

18. It has been reported that 83% of federal government employees use e-mail. If a sample of 200 federal government employees is selected, find the mean, variance, and standard deviation of the number who use e-mail.

Source: *USA TODAY.*

19. A survey found that 21% of Americans watch fireworks on television on July 4. Find the mean, variance, and standard deviation of the number of individuals who watch fireworks on television on July 4 if a random sample of 1000 Americans is selected.

Source: USA Snapshot, *USA TODAY.*

20. In a restaurant, a study found that 42% of all patrons smoked. If the seating capacity of the restaurant is 80 people, find the mean, variance, and standard deviation of the number of smokers. About how many seats should be available for smoking customers?

21. A survey found that 25% of pet owners had their pets bathed professionally rather than do it themselves. If 18 pet owners are randomly selected, find the probability that exactly five people have their pets bathed professionally.

Source: USA Snapshot, *USA TODAY.*

22. In a survey, 63% of Americans said they own an answering machine. If 14 Americans are selected at random, find the probability that exactly 9 own an answering machine.

Source: USA Snapshot, *USA TODAY.*

23. One out of every three Americans believes that the U.S. government should take “primary responsibility” for eliminating poverty in the United States. If 10 Americans are selected, find the probability that at most 3 will believe that the U.S. government should take primary responsibility for eliminating poverty.

Source: *Harper's Index.*

24. In a survey, 58% of American adults said they had never heard of the Internet. If 20 American adults are selected at random, find the probability that exactly 12 will say they have never heard of the Internet.

Source: *Harper's Index.*

25. In the past year, 13% of businesses have eliminated jobs. If five businesses are selected at random, find the probability that at least three have eliminated jobs during the last year.

Source: *USA TODAY.*

26. Of graduating high school seniors, 14% said that their generation will be remembered for their social concerns. If seven graduating seniors are selected at random, find the probability that either two or three will agree with that statement.

Source: *USA TODAY.*

27. A survey found that 86% of Americans have never been a victim of violent crime. If a sample of 12 Americans is selected at random, find the probability that 10 or more have never been victims of violent crime. Does it seem reasonable that 10 or more have never been victims of violent crime?

Source: *Harper's Index.*

Extending the Concepts

28. The graph shown here represents the probability distribution for the number of girls in a family of three children. From this graph, construct a probability distribution.

29. Construct a binomial distribution graph for the number of defective computer chips in a lot of four if $p = 0.3$.

Technology *Step by Step*

MINITAB
Step by Step

The Binomial Distribution

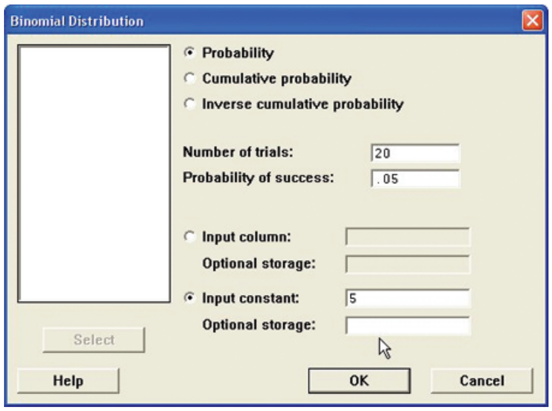
Calculate a Binomial Probability

From Example 5–19, it is known that 5% of the population is afraid of being alone at night. If a random sample of 20 Americans is selected, what is the probability that exactly 5 of them are afraid?

$$n = 20 \qquad p = 0.05\ (5\%) \qquad \text{and} \qquad X = 5\ (5 \text{ out of } 20)$$

No data need to be entered in the worksheet.

1. Select **Calc>Probability Distributions>Binomial.**
2. Click the option for Probability.
3. Click in the text box for Number of trials:.
4. Type in **20,** then Tab to Probability of success, then type **.05.**
5. Click the option for Input constant, then type in **5.** Leave the text box for Optional storage empty. If the name of a constant such as K1 is entered here, the results are stored but not displayed in the session window.



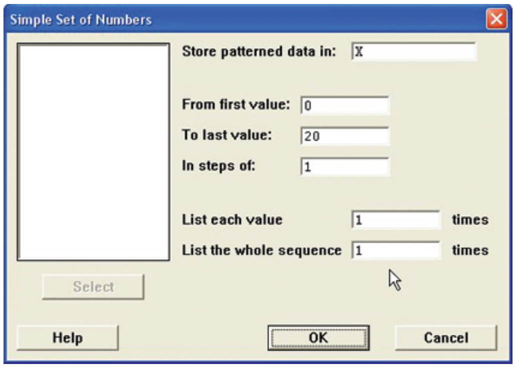
6. Click [OK]. The results are visible in the session window.

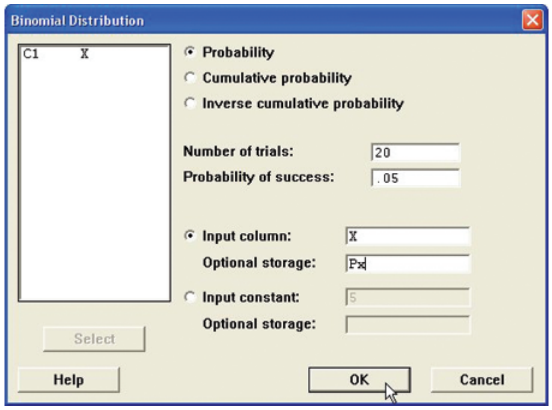
Probability Density Function
Binomial with n = 20 and p = 0.05
x f(x)
5 0.0022446

Construct a Binomial Distribution

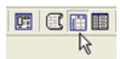
These instructions will use $n = 20$ and $p = 0.05$.

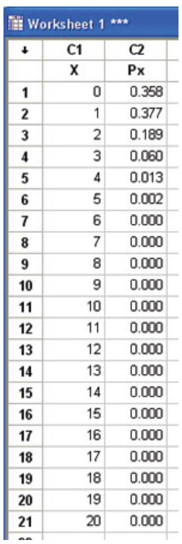
1. Select **Calc>Make Patterned Data>Simple Set of Numbers.**
2. You must enter three items:
   a) Enter **X** in the box for Store patterned data in:. MINITAB will use the first empty column of the active worksheet and name it X.
   b) Press Tab. Enter the value of **0** for the first value. Press Tab.
   c) Enter **20** for the last value. This value should be *n*. In steps of:, the value should be 1.
3. Click [OK].
4. Select **Calc>Probability Distributions>Binomial.**
5. In the dialog box you must enter five items.
   a) Click the button for Probability.
   b) In the box for Number of trials enter **20.**
   c) Enter **.05** in the Probability of success.





   d) Check the button for Input columns then type the column name, **X,** in the text box.
   e) Click in the box for Optional storage, then type **Px.**
6. Click [OK]. The first available column will be named Px, and the calculated probabilities will be stored in it.
7. To view the completed table, click the worksheet icon on the toolbar.



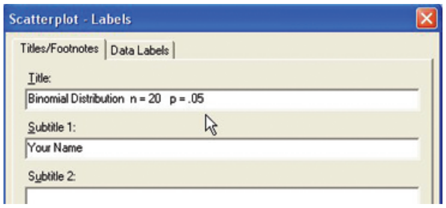


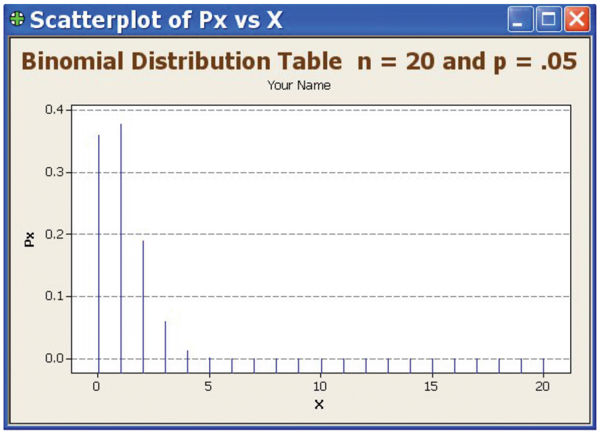
↓	C1	C2
	X	Px
1	0	0.358
2	1	0.377
3	2	0.189
4	3	0.060
5	4	0.013
6	5	0.002
7	6	0.000
8	7	0.000
9	8	0.000
10	9	0.000
11	10	0.000
12	11	0.000
13	12	0.000
14	13	0.000
15	14	0.000
16	15	0.000
17	16	0.000
18	17	0.000
19	18	0.000
20	19	0.000
21	20	0.000

Graph a Binomial Distribution

The table must be available in the worksheet.

1. Select **Graph>Scatterplot,** then Simple.
   a) Double-click on C2 Px for the Y variable and C1 X for the X variable.
   b) Click [Data view], then Project lines, then [OK]. Deselect any other type of display that may be selected in this list.
   c) Click on [Labels], then Title/Footnotes.
   d) Type an appropriate title, such as **Binomial Distribution n = 20, p = .05.**
   e) Press Tab to the Subtitle 1, then type in Your Name.
   f) Optional: Click [Scales] then [Gridlines] then check the box for Y major ticks.
   g) Click [OK] twice.





The graph will be displayed in a window. Right-click the control box to save, print, or close the graph.

TI-83 Plus or TI-84 Plus Step by Step

Binomial Random Variables

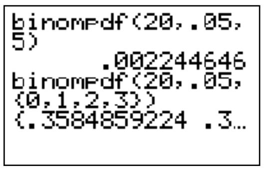
To find the probability for a binomial variable:
Press **2nd [DISTR]** then **0** for binomial pdf(
The form is binompdf(n,p,X).

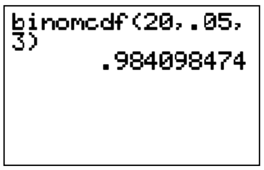
Example: $n = 20$, $X = 5$, $p = .05$. (Example 5–19*a* from the text) binompdf(20,.05,5)

Example: $n = 20$, $X = 0, 1, 2, 3$, $p = .05$. (Example 5–19*b* from the text)
binompdf(20,.05,{0,1,2,3})
The calculator will display the probabilities in a list. Use the arrow keys to view entire display.

To find the cumulative probability for a binomial random variable:
Press **2nd [DISTR]** then **A (ALPHA MATH)** for binomcdf(
The form is binomcdf(n,p,X). This will calculate the cumulative probability for values from 0 to X.

Example: $n = 20$, $X = 0, 1, 2, 3$, $p = .05$ (Example 5–19*b* from the text)
binomcdf(20,.05,3)

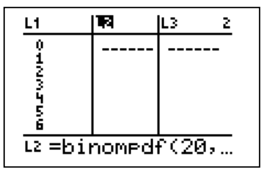


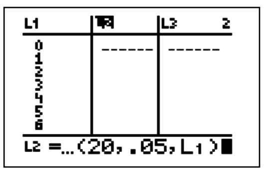


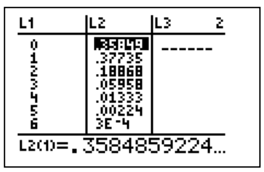
To construct a binomial probability table:

1. Enter the X values 0 through n into L_1.
2. Move the cursor to the top of the L_2 column so that L_2 is highlighted.
3. Type the command binompdf(n,p,L_1), then press **ENTER.**

Example: $n = 20$, $p = .05$ (Example 5–19 from the text)







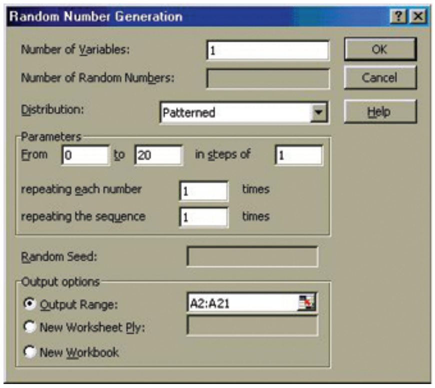
Excel
Step by Step

Creating a Binomial Distribution and Graph

These instructions will show how Excel can be used to construct a binomial distribution table, using $n = 20$ and $p = 0.35$.

1. Label cell A1 **X,** for the value of the random variable in a new worksheet.
2. Label cell B1 **P(X),** for the corresponding probabilities.
3. To enter the integers from 0 to 20 in column A starting at cell A2, select **Tools>Data Analysis>Random Number Generation,** then click OK.
4. In the Random Number Generation dialog box, enter the following:
   a) Number of Variables: **1**
   b) Distribution: Patterned
   c) Parameters: From **0** to **20** in steps of **1,** repeating each number: **1** times and repeating each sequence: **1** times
   d) Output range: **A2:A21**
5. Then click [OK].

Random Number Generation Dialog Box



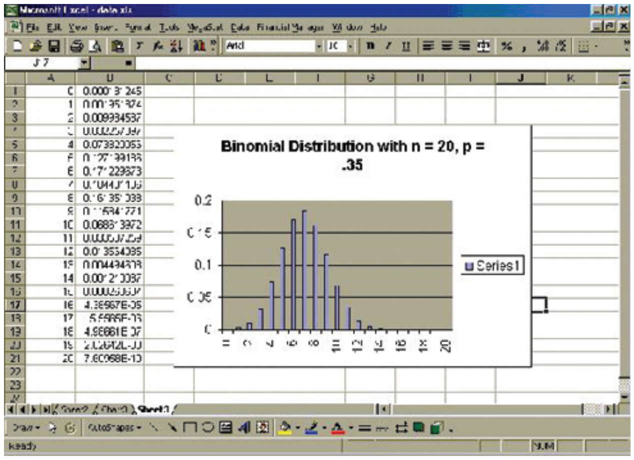
6. To create the probability corresponding to the first value of the binomial random variable, select cell B2 and type: **=BINOMDIST(0,20,.35,FALSE).** This will give the probability of obtaining 0 successes in 20 trials of a binomial experiment in which the probability of success is 0.35.
7. Repeat this procedure for each of the values of the random variable from column A.

Note: If you wish to obtain a column of cumulative probabilities for each outcome, you can type: **=BINOMDIST(0,20,.35,TRUE)** and repeat for each outcome. You can label the cell at the top of the column of cumulative probabilities, cell C1, as **P(X<=x).**

To create the graph:

1. Select the Chart Wizard from the Toolbar. Select the Column type and the first Chart sub-type. Click Next.
2. Click the icon next to the Data Range box to minimize the dialog box. Highlight the cells B1:B21, and then click the icon to return to the dialog box.
3. Select the Series tab. Click the icon next to the Category (X) axis labels box. From the worksheet, highlight cells A1:A21. Click Next. Label as necessary, then click Next and Finish.

The distribution and the graph are shown.



5–5 Other Types of Distributions (Optional)

In addition to the binomial distribution, other types of distributions are used in statistics. Three of the most commonly used distributions are the multinomial distribution, the Poisson distribution, and the hypergeometric distribution. They are described next.

The Multinomial Distribution

Objective 5

Find probabilities for outcomes of variables, using the Poisson, hypergeometric, and multinomial distributions.

Recall that in order for an experiment to be binomial, two outcomes are required for each trial. But if each trial in an experiment has more than two outcomes, a distribution called the **multinomial distribution** must be used. For example, a survey might require the responses of "approve," "disapprove," or "no opinion." In another situation, a person may have a choice of one of five activities for Friday night, such as a movie, dinner, baseball game, play, or party. Since these situations have more than two possible outcomes for each trial, the binomial distribution cannot be used to compute probabilities.

The multinomial distribution can be used for such situations if the probabilities for each trial remain constant and the outcomes are independent for a fixed number of trials. The events must also be mutually exclusive.

Formula for the Multinomial Distribution

If X consists of events $E_1, E_2, E_3, \ldots, E_k$, which have corresponding probabilities $p_1, p_2, p_3, \ldots, p_k$ of occurring, and X_1 is the number of times E_1 will occur, X_2 is the number of times E_2 will occur, X_3 is the number of times E_3 will occur, etc., then the probability that X will occur is

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot X_3! \cdots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdots p_k^{X_k}$$

where $X_1 + X_2 + X_3 + \cdots + X_k = n$ and $p_1 + p_2 + p_3 + \cdots + p_k = 1$.

Example 5–24

In a large city, 50% of the people choose a movie, 30% choose dinner and a play, and 20% choose shopping as a leisure activity. If a sample of five people is randomly selected, find the probability that three are planning to go to a movie, one to a play, and one to a shopping mall.

Solution

We know that $n = 5$, $X_1 = 3$, $X_2 = 1$, $X_3 = 1$, $p_1 = 0.50$, $p_2 = 0.30$, and $p_3 = 0.20$. Substituting in the formula gives

$$P(X) = \frac{5!}{3! \cdot 1! \cdot 1!} \cdot (0.50)^3(0.30)^1(0.20)^1 = 0.15$$

Again, note that the multinomial distribution can be used even though replacement is not done, provided that the sample is small in comparison with the population.

Example 5–25

In a music store, a manager found that the probabilities that a person buys zero, one, or two or more CDs are 0.3, 0.6, and 0.1, respectively. If six customers enter the store, find the probability that one won't buy any CDs, three will buy one CD, and two will buy two or more CDs.

Solution

It is given that $n = 6$, $X_1 = 1$, $X_2 = 3$, $X_3 = 2$, $p_1 = 0.3$, $p_2 = 0.6$, and $p_3 = 0.1$. Then

$$\begin{aligned} P(X) &= \frac{6!}{1!3!2!} \cdot (0.3)^1(0.6)^3(0.1)^2 \\ &= 60 \cdot (0.3)(0.216)(0.01) = 0.03888 \end{aligned}$$

Example 5–26

A box contains four white balls, three red balls, and three blue balls. A ball is selected at random, and its color is written down. It is replaced each time. Find the probability that if five balls are selected, two are white, two are red, and one is blue.

Solution

We know that $n = 5$, $X_1 = 2$, $X_2 = 2$, $X_3 = 1$; $p_1 = \frac{4}{10}$, $p_2 = \frac{3}{10}$, and $p_3 = \frac{3}{10}$; hence,

$$P(X) = \frac{5!}{2!2!1!} \cdot \left(\frac{4}{10}\right)^2\left(\frac{3}{10}\right)^2\left(\frac{3}{10}\right)^1 = \frac{81}{625}$$

Historical Notes

Simeon D. Poisson (1781–1840) formulated the distribution that bears his name. It appears only once in his writings and is only one page long. Mathematicians paid little attention to it until 1907, when a statistician named W. S. Gosset found real applications for it.

Thus, the multinomial distribution is similar to the binomial distribution but has the advantage of allowing one to compute probabilities when there are more than two outcomes for each trial in the experiment. That is, the multinomial distribution is a general distribution, and the binomial distribution is a special case of the multinomial distribution.

The Poisson Distribution

A discrete probability distribution that is useful when n is large and p is small and when the independent variables occur over a period of time is called the **Poisson distribution.** In addition to being used for the stated conditions (i.e., n is large, p is small, and the variables occur over a period of time), the Poisson distribution can be used when a density of items is distributed over a given area or volume, such as the number of plants growing per acre or the number of defects in a given length of videotape.

Formula for the Poisson Distribution

The probability of X occurrences in an interval of time, volume, area, etc., for a variable where λ (Greek letter lambda) is the mean number of occurrences per unit (time, volume, area, etc.) is

$$P(X;\lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad \text{where } X = 0, 1, 2, \ldots$$

The letter e is a constant approximately equal to 2.7183.

Round the answers to four decimal places.

Example 5–27

If there are 200 typographical errors randomly distributed in a 500-page manuscript, find the probability that a given page contains exactly three errors.

Solution

First, find the mean number λ of errors. Since there are 200 errors distributed over 500 pages, each page has an average of

$$\lambda = \frac{200}{500} = \frac{2}{5} = 0.4$$

or 0.4 error per page. Since $X = 3$, substituting into the formula yields

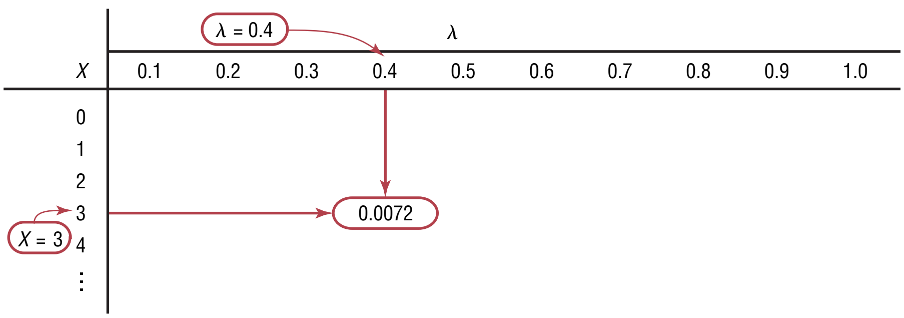
$$P(X;\lambda) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{(2.7183)^{-0.4}(0.4)^3}{3!} = 0.0072$$

Thus, there is less than a 1% probability that any given page will contain exactly three errors.

Since the mathematics involved in computing Poisson probabilities is somewhat complicated, tables have been compiled for these probabilities. Table C in Appendix C gives P for various values for λ and X.

In Example 5–27, where X is 3 and λ is 0.4, the table gives the value 0.0072 for the probability. See Figure 5–4.

Figure 5–4
Using Table C



Example 5–28

A sales firm receives, on the average, three calls per hour on its toll-free number. For any given hour, find the probability that it will receive the following.

a. At most three calls   *b.* At least three calls   *c.* Five or more calls

Solution

a. "At most three calls" means 0, 1, 2, or 3 calls. Hence,

$$P(0; 3) + P(1; 3) + P(2; 3) + P(3; 3)$$
$$= 0.0498 + 0.1494 + 0.2240 + 0.2240$$
$$= 0.6472$$

b. "At least three calls" means 3 or more calls. It is easier to find the probability of 0, 1, and 2 calls and then subtract this answer from 1 to get the probability of at least 3 calls.

$$P(0; 3) + P(1; 3) + P(2; 3) = 0.0498 + 0.1494 + 0.2240 = 0.4232$$

and

$$1 - 0.4232 = 0.5768$$

c. For the probability of five or more calls, it is easier to find the probability of getting 0, 1, 2, 3, or 4 calls and subtract this answer from 1. Hence,

$$P(0; 3) + P(1; 3) + P(2; 3) + P(3; 3) + P(4; 3)$$
$$= 0.0498 + 0.1494 + 0.2240 + 0.2240 + 0.1680$$
$$= 0.8152$$

and

$$1 - 0.8152 = 0.1848$$

Thus, for the events described, the part *a* event is most likely to occur and the part *c* event is least likely to occur.

The Poisson distribution can also be used to approximate the binomial distribution when the expected value $\lambda = n \cdot p$ is less than 5, as shown in Example 5–29. (The same is true when $n \cdot q < 5$.)

Example 5–29

If approximately 2% of the people in a room of 200 people are left-handed, find the probability that exactly five people there are left-handed.

Solution

Since $\lambda = n \cdot p$, then $\lambda = (200)(0.02) = 4$. Hence,

$$P(X; \lambda) = \frac{(2.7183)^{-4}(4)^5}{5!} = 0.1563$$

which is verified by the formula ${}_{200}C_5(0.02)^5(0.98)^{195} \approx 0.1579$. The difference between the two answers is based on the fact that the Poisson distribution is an approximation and rounding has been used.

The Hypergeometric Distribution

When sampling is done *without* replacement, the binomial distribution does not give exact probabilities, since the trials are not independent. The smaller the size of the population, the less accurate the binomial probabilities will be.

For example, suppose a committee of four people is to be selected from seven women and five men. What is the probability that the committee will consist of three women and one man?

To solve this problem, one must find the number of ways a committee of three women and one man can be selected from seven women and five men. This answer can be found by using combinations; it is

$$_{7}C_{3} \cdot {_{5}C_{1}} = 35 \cdot 5 = 175$$

Next, find the total number of ways a committee of 4 people can be selected from 12 people. Again, by the use of combinations, the answer is

$$_{12}C_{4} = 495$$

Finally, the probability of getting a committee of three women and one man from seven women and five men is

$$P(X) = \frac{175}{495} = \frac{35}{99}$$

The results of the problem can be generalized by using a special probability distribution called the hypergeometric distribution. The **hypergeometric distribution** is a distribution of a variable that has two outcomes when sampling is done without replacement.

The probabilities for the hypergeometric distribution can be calculated by using the formula given next.

Formula for the Hypergeometric Distribution

Given a population with only two types of objects (females and males, defective and nondefective, successes or failures, etc.), such that there are a items of one kind and b items of another kind and $a + b$ equals the total population, the probability $P(X)$ of selecting without replacement a sample of size n with X items of type a and $n - X$ items of type b is

$$P(X) = \frac{{_{a}C_{X}} \cdot {_{b}C_{n-X}}}{{_{a+b}C_{n}}}$$

The basis of the formula is that there are $_{a}C_{X}$ ways of selecting the first type of items, $_{b}C_{n-X}$ ways of selecting the second type of items, and $_{a+b}C_{n}$ ways of selecting n items from the entire population.

Example 5–30

Ten people apply for a job as assistant manager of a restaurant. Five have completed college and five have not. If the manager selects three applicants at random, find the probability that all three are college graduates.

Solution

Assigning the values to the variables gives

$a = 5$ college graduates $\quad n = 3$

$b = 5$ nongraduates $\quad X = 3$

and $n - X = 0$. Substituting in the formula gives

$$P(X) = \frac{{_{5}C_{3}} \cdot {_{5}C_{0}}}{{_{10}C_{3}}} = \frac{10}{120} = \frac{1}{12}$$

Example 5–31

A recent study found that four out of nine houses were underinsured. If five houses are selected from the nine houses, find the probability that exactly two are underinsured.

Solution

In this problem

$$a = 4 \qquad b = 5 \qquad n = 5 \qquad X = 2 \qquad n - X = 3$$

Then

$$P(X) = \frac{{}_4C_2 \cdot {}_5C_3}{{}_9C_5} = \frac{60}{126} = \frac{10}{21}$$

In many situations where objects are manufactured and shipped to a company, the company selects a few items and tests them to see whether they are satisfactory or defective. If a certain percentage is defective, the company then can refuse the whole shipment. This procedure saves the time and cost of testing every single item. To make the judgment about whether to accept or reject the whole shipment based on a small sample of tests, the company must know the probability of getting a specific number of defective items. To calculate the probability, the company uses the hypergeometric distribution.

Example 5–32

A lot of 12 compressor tanks is checked to see whether there are any defective tanks. Three tanks are checked for leaks. If one or more of the three is defective, the lot is rejected. Find the probability that the lot will be rejected if there are actually three defective tanks in the lot.

Solution

Since the lot is rejected if at least one tank is found to be defective, it is necessary to find the probability that none are defective and subtract this probability from 1.

Here, $a = 3$, $b = 9$, $n = 3$, $X = 0$; so

$$P(X) = \frac{{}_3C_0 \cdot {}_9C_3}{{}_{12}C_3} = \frac{1 \cdot 84}{220} = 0.38$$

Hence,

$$P(\text{at least one defective}) = 1 - P(\text{no defectives}) = 1 - 0.38 = 0.62$$

There is a 0.62, or 62%, probability that the lot will be rejected when 3 of the 12 tanks are defective.

A summary of the discrete distributions used in this chapter is shown in Table 5–1.

Interesting Fact

An IBM supercomputer set a world record in 2004 by performing 36.01 trillion calculations in 1 second.

Table 5–1 Summary of Discrete Distributions

1. Binomial distribution

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

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

Used when there are only two independent outcomes for a fixed number of independent trials and the probability for each success remains the same for each trial.

2. Multinomial distribution

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot X_3! \cdots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdots p_k^{X_k}$$

where

$$X_1 + X_2 + X_3 + \cdots + X_k = n \qquad \text{and} \qquad p_1 + p_2 + p_3 + \cdots + p_k = 1$$

Used when the distribution has more than two outcomes, the probabilities for each trial remain constant, outcomes are independent, and there are a fixed number of trials.

3. Poisson distribution

$$P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad \text{where } X = 0, 1, 2, \ldots$$

Used when n is large and p is small, the independent variable occurs over a period of time, or a density of items is distributed over a given area or volume.

4. Hypergeometric distribution

$$P(X) = \frac{{}_aC_X \cdot {}_bC_{n-X}}{{}_{a+b}C_n}$$

Used when there are two outcomes and sampling is done without replacement.

Applying the Concepts 5–5

Rockets and Targets

During the latter days of World War II, the Germans developed flying rocket bombs. These bombs were used to attack London. Allied military intelligence didn't know whether these bombs were fired at random or had a sophisticated aiming device. To determine the answer, they used the Poisson distribution.

To assess the accuracy of these bombs, London was divided into 576 square regions. Each region was $\frac{1}{4}$ square kilometer in area. They then compared the number of actual hits with the theoretical number of hits by using the Poisson distribution. If the values in both distributions were close, then they would conclude that the rockets were fired at random. The actual distribution is as follows:

Hits	0	1	2	3	4	5
Regions	229	211	93	35	7	1

1. Using the Poisson distribution, find the theoretical values for each number of hits. In this case, the number of bombs was 535, and the number of regions was 576. So

$$\mu = \frac{535}{576} = 0.929$$

For three hits,

$$P(X) = \frac{\mu^X \cdot e^{-\mu}}{X!}$$
$$= \frac{(0.929)^3(2.7183)^{-0.929}}{3!} = 0.0528$$

Hence the number of hits is (0.0528)(576) = 30.4128.

Complete the table for the other number of hits.

Hits	0	1	2	3	4	5
Regions				30.4		

2. Write a brief statement comparing the two distributions.
3. Based on your answer to question 2, can you conclude that the rockets were fired at random?

See page 284 for the answer.

Exercises 5–5

1. Use the multinomial formula and find the probabilities for each.

a. $n = 6, X_1 = 3, X_2 = 2, X_3 = 1, p_1 = 0.5, p_2 = 0.3, p_3 = 0.2$

b. $n = 5, X_1 = 1, X_2 = 2, X_3 = 2, p_1 = 0.3, p_2 = 0.6, p_3 = 0.1$

c. $n = 4, X_1 = 1, X_2 = 1, X_3 = 2, p_1 = 0.8, p_2 = 0.1, p_3 = 0.1$

d. $n = 3, X_1 = 1, X_2 = 1, X_3 = 1, p_1 = 0.5, p_2 = 0.3, p_3 = 0.2$

e. $n = 5, X_1 = 1, X_2 = 3, X_3 = 1, p_1 = 0.7, p_2 = 0.2, p_3 = 0.1$

2. The probabilities that a textbook page will have 0, 1, 2, or 3 typographical errors are 0.79, 0.12, 0.07, and 0.02, respectively. If eight pages are randomly selected, find the probability that four will contain no errors, two will contain 1 error, one will contain 2 errors, and one will contain 3 errors.

3. The probabilities are 0.25, 0.40, and 0.35 that an 18-wheel truck will have 0 violations, 1 violation, or 2 or more violations when it is given a safety inspection. If eight trucks are inspected, find the probability that three will have 0 violations, two will have 1 violation, and three will have 2 or more violations.

4. When a customer enters a pharmacy, the probabilities that he or she will have 0, 1, 2, or 3 prescriptions filled are 0.60, 0.25, 0.10, and 0.05, respectively. For a sample of six people who enter the pharmacy, find the probability that two will have 0 prescriptions, two will have 1 prescription, one will have 2 prescriptions, and one will have 3 prescriptions.

5. A die is rolled 4 times. Find the probability of two 1s, one 2, and one 3.

6. According to Mendel's theory, if tall and colorful plants are crossed with short and colorless plants, the corresponding probabilities are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$ for tall and colorful, tall and colorless, short and colorful, and short and colorless, respectively. If eight plants are selected, find the probability that one will be tall and colorful, three will be tall and colorless, three will be short and colorful, and one will be short and colorless.

7. Find each probability $P(X; \lambda)$, using Table C in Appendix C.

a. $P(5; 4)$

b. $P(2; 4)$

c. $P(6; 3)$

d. $P(10; 7)$

e. $P(9; 8)$

8. If 2% of the batteries manufactured by a company are defective, find the probability that in a case of 144 batteries, there are 3 defective ones.

9. A recent study of robberies for a certain geographic region showed an average of one robbery per 20,000 people. In a city of 80,000 people, find the probability of the following.

a. No robberies

b. One robbery

c. Two robberies

d. Three or more robberies

10. In a 400-page manuscript, there are 200 randomly distributed misprints. If a page is selected, find the probability that it has one misprint.

11. A telephone soliciting company obtains an average of five orders per 1000 solicitations. If the company reaches 250 potential customers, find the probability of obtaining at least two orders.

12. A mail-order company receives an average of five orders per 500 solicitations. If it sends out 100 advertisements, find the probability of receiving at least two orders.

13. A videotape has an average of one defect every 1000 feet. Find the probability of at least one defect in 3000 feet.

14. If 3% of all cars fail the emissions inspection, find the probability that in a sample of 90 cars, 3 will fail. Use the Poisson approximation.

15. The average number of phone inquiries per day at the poison control center is four. Find the probability it will receive five calls on a given day. Use the Poisson approximation.

16. In a batch of 2000 calculators, there are, on average, eight defective ones. If a random sample of 150 is selected, find the probability of five defective ones.

17. In a camping club of 18 members, nine prefer hoods and nine prefer hats and earmuffs. On a recent winter outing attended by six members, find the probability that exactly three members wore earmuffs and hats.

18. A bookstore owner examines 5 books from each lot of 25 to check for missing pages. If he finds at least two books with missing pages, the entire lot is returned. If, indeed, there are five books with missing pages, find the probability that the lot will be returned.

19. Shirts are packed at random in two sizes, regular and extra large. Four shirts are selected from a box of 24 and checked for size. If there are 15 regular shirts in the box, find the probability that all 4 will be regular size.

20. A shipment of 24 computer keyboards is rejected if 4 are checked for defects and at least 1 is found to be defective. Find the probability that the shipment will be returned if there are actually 6 defective keyboards.

21. A shipment of 24 electric typewriters is rejected if 3 are checked for defects and at least 1 is found to be defective. Find the probability that the shipment will be returned if there are actually 6 typewriters that are defective.

Technology *Step by Step*

TI-83 Plus or TI-84 Plus Step by Step

Poisson Random Variables

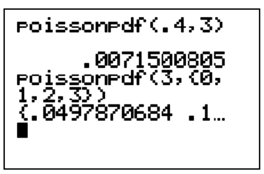
To find the probability for a Poisson random variable:
Press **2nd [DISTR]** then **B (ALPHA APPS)** for poissonpdf(
The form is poissonpdf(λ,X).

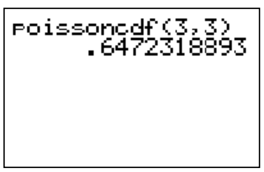
Example: $\lambda = 0.4$, $X = 3$ (Example 5–27 from the text)
poissonpdf(.4,3)

Example: $\lambda = 3$, $X = 0, 1, 2, 3$ (Example 5–28*a* from the text)
poissonpdf(3,{0,1,2,3})
The calculator will display the probabilities in a list. Use the arrow keys to view the entire display.

To find the cumulative probability for a Poisson random variable:
Press **2nd [DISTR]** then **C (ALPHA PRGM)** for poissoncdf(
The form is poissoncdf(λ,X). This will calculate the cumulative probability for values from 0 to X.

Example: $\lambda = 3$, $X = 0, 1, 2, 3$ (Example 5–28*a* from the text)
poissoncdf(3,3)

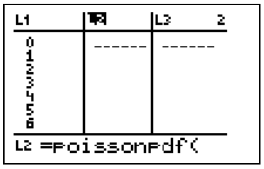


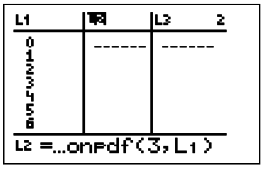


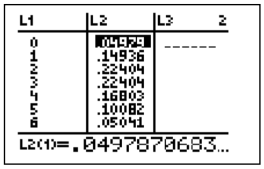
To construct a Poisson probability table:

1. Enter the X values 0 through a large possible value of X into L_1.
2. Move the cursor to the top of the L_2 column so that L_2 is highlighted.
3. Enter the command poissonpdf(λ,L_1) then press **ENTER.**

Example: $\lambda = 3$, $X = 0, 1, 2, 3, \ldots, 10$ (Example 5–28 from the text)







5–6 Summary

Many variables have special probability distributions. This chapter presented several of the most common probability distributions, including the binomial distribution, the multinomial distribution, the Poisson distribution, and the hypergeometric distribution.

The binomial distribution is used when there are only two outcomes for an experiment, there are a fixed number of trials, the probability is the same for each trial, and the outcomes are independent of one another. The multinomial distribution is an extension of the binomial distribution and is used when there are three or more outcomes for an experiment. The hypergeometric distribution is used when sampling is done without replacement. Finally, the Poisson distribution is used in special cases when independent events occur over a period of time, area, or volume.

A probability distribution can be graphed, and the mean, variance, and standard deviation can be found. The mathematical expectation can also be calculated for a probability distribution. Expectation is used in insurance and games of chance.

Important Terms

Important Formulas

Formula for the mean of a probability distribution:

$$\mu = \Sigma X \cdot P(X)$$

Formulas for the variance and standard deviation of a probability distribution:

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

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

Formula for expected value:

$$E(X) = \Sigma X \cdot P(X)$$

Binomial probability formula:

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

Formula for the mean of the binomial distribution:

$$\mu = n \cdot p$$

Formulas for the variance and standard deviation of the binomial distribution:

$$\sigma^2 = n \cdot p \cdot q \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

Formula for the multinomial distribution:

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot X_3! \cdot \cdot \cdot X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdot \cdot \cdot p_k^{X_k}$$

Formula for the Poisson distribution:

$$P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad \text{where } X = 0, 1, 2, \ldots$$

Formula for the hypergeometric distribution:

$$P(X) = \frac{{}_aC_X \cdot {}_bC_{n-X}}{{}_{a+b}C_n}$$

Review Exercises

For Exercises 1 through 3, determine whether the distribution represents a probability distribution. If it does not, state why.

1.

X	1	2	3	4	5
$P(X)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

2.

X	10	20	30
$P(X)$	0.1	0.4	0.3

3.

X	8	12	16	20
$P(X)$	$\frac{5}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

4. The number of emergency calls a local police department receives per 24-hour period is distributed as shown here. Construct a graph for the data.

Number of calls X	10	11	12	13	14
Probability $P(X)$	0.02	0.12	0.40	0.31	0.15

5. A study was conducted to determine the number of radios each household has. The data are shown here. Draw a graph for the data.

Number of radios	0	1	2	3	4
Probability $P(X)$	0.05	0.30	0.45	0.12	0.08

6. A box contains five pennies, three dimes, one quarter, and one half-dollar. Construct a probability distribution and draw a graph for the data.

7. At Tyler's Tie Shop, Tyler found the probabilities that a customer will buy 0, 1, 2, 3, or 4 ties, as shown. Construct a graph for the distribution.

Number of ties X	0	1	2	3	4
Probability $P(X)$	0.30	0.50	0.10	0.08	0.02

8. A bank has a drive-through service. The number of customers arriving during a 15-minute period is distributed as shown. Find the mean, variance, and standard deviation for the distribution.

Number of customers X	0	1	2	3	4
Probability $P(X)$	0.12	0.20	0.31	0.25	0.12

9. At a small community museum, the number of visitors per hour during the day has the distribution shown here. Find the mean, variance, and standard deviation for the data.

Number of visitors X	13	14	15	16	17
Probability $P(X)$	0.12	0.15	0.29	0.25	0.19

10. During a recent paint sale at Corner Hardware, the number of cans of paint purchased was distributed as shown. Find the mean, variance, and standard deviation of the distribution.

Number of cans X	1	2	3	4	5
Probability $P(X)$	0.42	0.27	0.15	0.10	0.06

11. The number of inquiries received per day for a college catalog is distributed as shown. Find the mean, variance, and standard deviation for the data.

Number of inquiries X	22	23	24	25	26	27
Probability $P(X)$	0.08	0.19	0.36	0.25	0.07	0.05

12. A producer plans an outdoor regatta for May 3. The cost of the regatta is $8000. This includes advertising, security, printing tickets, entertainment, etc. The producer plans to make $15,000 profit if all goes well. However, if it rains, the regatta would have to be canceled. According to the weather report, the

probability of rain is 0.3. Find the producer's expected profit.

13. A person selects a card from a deck. If it is a red card, he wins $1. If it is a black card between or including 2 and 10, he wins $5. If it is a black face card, he wins $10; and if it is a black ace, he wins $100. Find the expectation of the game. How much should a person bet if the game is to be fair?

14. If 30% of all commuters ride the train to work, find the probability that if 10 workers are selected, 5 will ride the train.

15. If 90% of all people between the ages of 30 and 50 drive a car, find these probabilities for a sample of 20 people in that age group.

a. Exactly 20 drive a car.
b. At least 15 drive a car.
c. At most 15 drive a car.

16. If 10% of the people who are given a certain drug experience dizziness, find these probabilities for a sample of 15 people who take the drug.

a. At least two people will become dizzy.
b. Exactly three people will become dizzy.
c. At most four people will become dizzy.

17. If 75% of nursing students are able to pass a drug calculation test, find the mean, variance, and standard deviation of the number of students who pass the test in a sample of 180 nursing students.

18. A club has 225 members. If there is a 70% attendance rate per meeting, find the mean, variance, and standard deviation of the number of people who will be present at each meeting.

19. The chance that a U.S. police chief believes the death penalty "significantly reduces the number of homicides" is 1 in 4. If a random sample of eight police chiefs is selected, find the probability that at most three believe that the death penalty significantly reduces the number of homicides.

Source: *Harper's Index.*

20. *American Energy Review* reported that 27% of American households burn wood. If a random sample of 500 American households is selected, find the mean, variance, and standard deviation of the number of households that burn wood.

Source: *100% American* by Daniel Evan Weiss.

21. Three out of four American adults under age 35 have eaten pizza for breakfast. If a random sample of 20 adults under age 35 is selected, find the probability that exactly 16 have eaten pizza for breakfast.

Source: *Harper's Index.*

22. One out of four Americans over age 55 has eaten pizza for breakfast. If a sample of 10 Americans over age 55 is selected at random, find the probability that at most 3 have eaten pizza for breakfast.

Source: *Harper's Index.*

23. **(Opt.)** The probabilities that a person will make 0, 1, 2, and 3 errors on an insurance claim are 0.70, 0.20, 0.08, and 0.02, respectively. If 20 claims are selected, find the probability that 12 will contain no errors, 4 will contain 1 error, 3 will contain 2 errors, and 1 will contain 3 errors.

24. **(Opt.)** Before a VCR leaves the factory, it is given a quality control check. The probabilities that a VCR contains 0, 1, or 2 defects are 0.90, 0.06, and 0.04, respectively. In a sample of 12 recorders, find the probability that 8 have 0 defects, 3 have 1 defect, and 1 has 2 defects.

25. **(Opt.)** In a Christmas display, the probability that all lights are the same color is 0.50; that 2 colors are used is 0.40; and that 3 or more colors are used is 0.10. If a sample of 10 displays is selected, find the probability that 5 have only 1 color of light, 3 have 2 colors, and 2 have 3 or more colors.

26. **(Opt.)** If 4% of the population carries a certain genetic trait, find the probability that in a sample of 100 people, there are exactly 8 people who have the trait. Assume the distribution is approximately Poisson.

27. **(Opt.)** Computer Help Hot Line receives, on the average, six calls per hour asking for assistance. The distribution is Poisson. For any randomly selected hour, find the probability that the company will receive

a. At least six calls.
b. Four or more calls.
c. At most five calls.

28. **(Opt.)** The number of boating accidents on Lake Emilie follows a Poisson distribution. The probability of an accident is 0.003. If there are 1000 boats on the lake during a summer month, find the probability that there will be 6 accidents.

29. **(Opt.)** If five cards are drawn from a deck, find the probability that two will be hearts.

30. **(Opt.)** Of the 50 automobiles in a used-car lot, 10 are white. If five automobiles are selected to be sold at an auction, find the probability that exactly two will be white.

31. **(Opt.)** A board of directors consists of seven men and five women. If a slate of three officers is selected, find these probabilities.

a. Exactly two are men.
b. All three are women.
c. Exactly two are women.

Statistics Today

Is Pooling Worthwhile?—Revisited

In the case of the pooled sample, the probability that only one test will be needed can be determined by using the binomial distribution. The question being asked is, In a sample of 15 individuals, what is the probability that no individual will have the disease? Hence, $n = 15$, $p = 0.05$, and $X = 0$. From Table B in Appendix C, the probability is 0.463, or 46% of the time, only one test will be needed. For screening purposes, then, pooling samples in this case would save considerable time, money, and effort as opposed to testing every individual in the population.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. The expected value of a random variable can be thought of as a long-run average.
2. The number of courses a student is taking this semester is an example of a continuous random variable.
3. When the multinomial distribution is used, the outcomes must be dependent.
4. A binomial experiment has a fixed number of trials.

Complete these statements with the best answer.

5. Random variable values are determined by ________.
6. The mean for a binomial variable can be found by using the formula ________.
7. One requirement for a probability distribution is that the sum of all the events in the sample space must equal ________.

Select the best answer.

8. What is the sum of the probabilities of all outcomes in a probability distribution?

   a. 0
   b. $\frac{1}{2}$
   c. 1
   d. It cannot be determined.

9. How many outcomes are there in a binomial experiment?

   a. 0
   b. 1
   c. 2
   d. It varies.

10. The number of plants growing in a specific area can be approximated by what distribution?

   a. Binomial
   b. Multinomial
   c. Hypergeometric
   d. Poisson

For questions 11 through 14, determine if the distribution represents a probability distribution. If not, state why.

11.

X	1	2	3	4	5
$P(X)$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

12.

X	3	6	9	12	15
$P(X)$	0.3	0.5	0.1	0.08	0.02

13.

X	50	75	100
$P(X)$	0.5	0.2	0.3

14.

X	4	8	12	16
$P(X)$	$\frac{1}{6}$	$\frac{3}{12}$	$\frac{1}{2}$	$\frac{1}{12}$

15. The number of fire calls the Conestoga Valley Fire Company receives per day is distributed as follows:

Number X	5	6	7	8	9
Probability $P(X)$	0.28	0.32	0.09	0.21	0.10

Construct a graph for the data.

16. A study was conducted to determine the number of telephones each household has. The data are shown here.

Number of telephones	0	1	2	3	4
Frequency	2	30	48	13	7

Construct a probability distribution and draw a graph for the data.

17. During a recent CD sale at Matt's Music Store, the number of CDs customers purchased was distributed as follows:

Number X	0	1	2	3	4
Probability $P(X)$	0.10	0.23	0.31	0.27	0.09

Find the mean, variance, and standard deviation of the distribution.

18. The number of calls received per day at a crisis hot line is distributed as follows:

Number X	30	31	32	33	34
Probability $P(X)$	0.05	0.21	0.38	0.25	0.11

Find the mean, variance, and standard deviation of the distribution.

19. There are six playing cards placed face down in a box. They are the 4 of diamonds, the 5 of hearts, the 2 of clubs, the 10 of spades, the 3 of diamonds, and the 7 of hearts. A person selects a card. Find the expected value of the draw.

20. A person selects a card from an ordinary deck of cards. If it is a black card, she wins $2. If it is a red card between or including 3 and 7, she wins $10. If it is a red face card, she wins $25; and if it is a black jack, she wins an extra $100. Find the expectation of the game.

21. If 40% of all commuters ride to work in carpools, find the probability that if eight workers are selected, five will ride in carpools.

22. If 60% of all women are employed outside the home, find the probability that in a sample of 20 women,

a. Exactly 15 are employed.
b. At least 10 are employed.
c. At most five are not employed outside the home.

23. If 80% of the applicants are able to pass a driver's proficiency road test, find the mean, variance, and standard deviation of the number of people who pass the test in a sample of 300 applicants.

24. A history class has 75 members. If there is a 12% absentee rate per class meeting, find the mean, variance, and standard deviation of the number of students who will be absent from each class.

25. The probability that a person will make zero, one, two, or three errors on his or her income tax return is 0.50, 0.30, 0.15, and 0.05, respectively. If 30 claims are selected, find the probability that 15 will contain 0 errors, 8 will contain one error, 5 will contain two errors, and 2 will contain three errors.

26. Before a television set leaves the factory, it is given a quality control check. The probability that a television contains zero, one, or two defects is 0.88, 0.08, and 0.04, respectively. In a sample of 16 televisions, find the probability that 9 will have no defects, 4 will have one defect, and 3 will have two defects.

27. Among the teams in a bowling league, the probability that the uniforms are all one color is 0.45, that two colors are used is 0.35, and that three or more colors are used is 0.20. If a sample of 12 uniforms is selected, find the probability that 5 contain only one color, 4 contain two colors, and 3 contain three or more colors.

28. If 8% of the population of trees are elm trees, find the probability that in a sample of 100 trees, there are exactly 6 elm trees. Assume the distribution is approximately Poisson.

29. Sports Scores Hot Line receives, on the average, eight calls per hour requesting the latest sports scores. The distribution is Poisson in nature. For any randomly selected hour, find the probability that the company will receive

a. At least eight calls.
b. Three or more calls.
c. At most seven calls.

30. There are 48 raincoats for sale at a local men's clothing store. Twelve are black. If six raincoats are selected to be marked down, find the probability that exactly three will be black.

31. A youth group has eight boys and six girls. If a slate of four officers is selected, find the probability that exactly

a. Three are girls.
b. Two are girls.
c. Four are boys.

Critical Thinking Challenges

1. Pennsylvania has a lottery entitled "Big 4." To win, a player must correctly match four digits from a daily lottery in which four digits are selected. Find the probability of winning.

2. In the Big 4 lottery, for a bet of $100, the payoff is $5000. What is the expected value of winning? Is it worth it?

3. If you played the same four-digit number every day (or any four-digit number for that matter) in the Big 4, how often (in years) would you win, assuming you have average luck?

4. In the game Chuck-a-Luck, three dice are rolled. A player bets a certain amount (say $1.00) on a number from 1 to 6. If the number appears on one die, the person wins $1.00. If it appears on two dice, the person wins $2.00, and if it appears on all three dice, the person wins $3.00. What are the chances of winning $1.00? $2.00? $3.00?

5. What is the expected value of the game of Chuck-a-Luck if a player bets $1.00 on one number?

Data Projects

Probability Distributions Roll three dice 100 times, recording the sum of the spots on the faces as you roll. Then find the average of the spots. How close is this to the theoretical average? Refer to Exercise 21 on page 254.

Answers to Applying the Concepts

Section 5–2 Dropping College Courses

1. The random variable under study is the reason for dropping a college course.
2. There were a total of 144 people in the study.
3. The complete table is as follows:

Reason for dropping a college course	Frequency	Percentage
Too difficult	45	31.25
Illness	40	27.78
Change in work schedule	20	13.89
Change of major	14	9.72
Family-related problems	9	6.25
Money	7	4.86
Miscellaneous	6	4.17
No meaningful reason	3	2.08

4. The probability that a student will drop a class because of illness is about 28%. The probability that a student will drop a class because of money is about 5%. The probability that a student will drop a class because of a change of major is about 10%.
5. The information is not itself a probability distribution, but it can be used as one.
6. The categories are not necessarily mutually exclusive, but we treated them as such in computing the probabilities.
7. The categories are not independent.
8. The categories are exhaustive.
9. Since all the probabilities are between 0 and 1, inclusive, and the probabilities sum to 1, the requirements for a discrete probability distribution are met.

Section 5–3 Expected Value

1. The expected value is the mean in a discrete probability distribution.
2. We would expect variation from the expected value of 3.
3. Answers will vary. One possible answer is that pregnant mothers in that area might be overly concerned upon hearing that the number of cases of kidney problems in newborns was nearly 4 times what was usually expected. Other mothers (particularly those who had taken a statistics course!) might ask for more information about the claim.
4. Answers will vary. One possible answer is that it does seem unlikely to have 11 newborns with kidney problems when we expect only 3 newborns to have kidney problems.
5. The public might better be informed by percentages or rates (e.g., rate per 1000 newborns).
6. The increase of eight babies born with kidney problems represents a 0.32% increase (less than $\frac{1}{2}$ of 1%).
7. Answers will vary. One possible answer is that the percentage increase does not seem to be something to be overly concerned about.

Section 5–4 Unsanitary Restaurants

1. The probability of eating at 3 restaurants with unsanitary conditions out of the 10 restaurants is 0.18651.
2. The probability of eating at 4 or 5 restaurants with unsanitary conditions out of the 10 restaurants is $0.24623 + 0.22291 = 0.46914$.
3. To find this probability, you could add the probabilities for eating at 1, 2, . . . , 10 unsanitary restaurants. An easier way to compute the probability is to subtract the probability of eating at no unsanitary restaurants from 1 (using the complement rule).
4. The highest probability for this distribution is 4, but the expected number of unsanitary restaurants that you would eat at is $10 \cdot \frac{3}{7} = 4.3$.
5. The standard deviation for this distribution is $\sqrt{10(\frac{3}{7})(\frac{4}{7})} = 1.56$.
6. This is a binomial distribution. We have two possible outcomes: "success" is eating in an unsanitary restaurant; "failure" is eating in a sanitary restaurant. The probability that one restaurant is unsanitary is independent of the probability that any other restaurant

is unsanitary. The probability that a restaurant is unsanitary remains constant at $\frac{3}{7}$. And we are looking at the number of unsanitary restaurants that we eat at out of 10 "trials."

7. The likelihood of success will vary from situation to situation. Just because we have two possible outcomes, this does not mean that each outcome occurs with probability 0.50.

Section 5–5 Rockets and Targets

1. The theoretical values for the number of hits are

Hits	0	1	2	3	4	5
Regions	227.5	211.3	98.2	30.4	7.1	1.3

2. The actual values are very close to the theoretical values.
3. Since the actual values are close to the theoretical values, it does appear that the rockets were fired at random.