
CHAPTER 6

PROBABILITY DISTRIBUTIONS FOR DISCRETE RANDOM VARIABLES

- 6.1 INTRODUCTION
 - 6.2 RANDOM VARIABLES
 - 6.3 DISCRETE PROBABILITY DISTRIBUTIONS
 - 6.4 MEAN AND VARIANCE
 - 6.5 THE BINOMIAL PROBABILITY DISTRIBUTION
 - 6.6 USING MINITAB (OPTIONAL)
 - 6.7 SUMMARY
- REVIEW EXERCISES

6.1

INTRODUCTION

In Chapters 2 and 3, we were concerned with the analysis of numerical data. In Chapter 5, we studied the concepts of experiment, outcome, sample space, and probability. In this chapter, we bring these ideas together by introducing **random variables**, which describe numerical properties of the elements of a sample space.

In this text, we will ultimately be interested in the study of populations. We will see that **probability distributions**, which are studied in this chapter and Chapter 7, can be used to summarize the essential features of a population of data values.

6.2

RANDOM VARIABLES

In any scientific inquiry, it is helpful to translate that which is being studied into numbers. For instance, if we are interested in the fuel efficiency of various kinds of automobiles, we would look at the mileage rating (number of miles per gallon of fuel) for each kind of car. The process of assigning numbers to the objects being studied seems to be a natural one.

Given a sample space, we often study some numerical property of the various outcomes in the sample space. For example, consider the collection of adult Americans. (This is the sample space associated with the experiment of selecting an adult American at random.) One possible numerical property of interest might be the age of each person. Once we have specified the numerical property of interest, we have in essence given a rule for assigning a certain number to each element of the sample space. (Next to the name of each adult American, we assign the number that represents that person's age.) A rule for assigning such numbers is called a **random variable**.

Definition A rule that enables us to assign a number to each outcome of a sample space is called a *random variable* on the sample space. The actual number associated with a particular outcome is called the value (or data value) of the random variable associated with this outcome. When the experiment is performed, the value associated with the outcome is said to have *occurred* or been observed.

EXAMPLE 6.1

In the 1992 presidential election, a pollster was assigned to interview the oldest male and female in various households. Each was asked whether he or she voted for George Bush. If the male said "Yes" and the female said "No," the pollster recorded YN. The interview of one household can be viewed as an experiment with sample space

$$S = \{YN, NY, YY, NN\}$$

Consider the random variable "number of yes responses" from the two voters. The possible values are 0, 1, and 2, as shown in the following table:

Outcome	YN	NY	YY	NN
Number of Yes Responses	1	1	2	0

EXAMPLE 6.2

Consider the collection (sample space) of all working people in the United States. We could consider a number of random variables, such as weight, height, age, salary, and so on.

We often use a symbol such as x , y , or z to represent an arbitrary or unspecified value of a random variable. Thus in Example 6.1, we could have used x to represent any one of the values 0, 1, or 2. We could also use the symbol more abstractly to represent the random variable itself. Thus we could say: Consider the random variable x , where x = the number of yes responses.

Suppose we let x = the water temperature at some randomly selected point in Lake Erie next September 1. Then x could potentially be any value between freezing (32°F) and boiling (212°F). (A spot containing industrial waste might account for the possibility of extremely high temperatures.)

$$32^\circ\text{F} < x < 212^\circ\text{F}$$

This inequality says that x must be greater than 32°F and less than 212°F . The numbers that satisfy this inequality constitute what is called an **interval** or **line segment**. The random variable “Water temperature” is a **continuous** random variable; it can potentially assume any value on an interval [Figure 6.1(a)].

On the other hand, suppose that we select three light bulbs from a production line. Let y be the random variable that counts the number of defectives. So y can potentially assume any one of the values 0, 1, 2, or 3. We call y a **discrete** random variable; the values it can potentially assume constitute separated or isolated points on the real number axis. For example, we could get 1 or 2 defectives but not 1.5 defectives. See Figure 6.1(b).

Definitions A random variable is *continuous* if it potentially can take on any value on some line segment or interval (that is, there are no “breaks” between possible values). A random variable is *discrete* if the values it can potentially assume constitute a sequence of isolated or separated points on the real number axis.

Continuous random variables usually measure the amount of something, whereas discrete random variables usually count something.

Here are some additional examples of continuous random variables: a person’s height, the length of time to run a marathon, the mass in kilograms of a celestial object (planet, star, meteor, piece of space dust, etc.). In this last example, we would consider the mass to be theoretically any value greater than 0, showing that sometimes the interval of possible values of a random variable is considered to be infinite in length.

Additional examples of discrete random variables are the number of children in a family, the number of times a person catches a cold in a given year, and the number of tosses of a coin before a tail appears. This last example shows that sometimes the sequence of potential values of a discrete random variable can be infinite because the sequence in this example would be $1, 2, 3, \dots$.

We will ultimately be interested in probabilities associated with various values of a random variable. A formula or table that enables us to find such probabilities is called a **probability distribution** for the random variable. We investigate probability distributions for discrete random variables in this chapter. In Chapter 7, we will study probability distributions for continuous random variables.

EXERCISES

- 6.1 Classify the following random variables as discrete or continuous:
 - (a) The number of automobile fatalities over a 72-hour period
 - (b) The length of time to run a 100-yard dash
 - (c) The diameter of pebbles from a stream
 - (d) The number of incoming phone calls to a switchboard at a city hospital
 - (e) The barometric pressure at noon at a specific location
 - (f) The number of bills paid on time per month to a plumber
- 6.2 One word is randomly selected from the following sentence: The grass needs cutting and should be cut by tomorrow afternoon. Let x be the number of letters in the selected word.
 - (a) What values can x assume?
 - (b) Is x a discrete or continuous random variable?
- 6.3 A green, a red, and a blue die are tossed. In each of the following, find the values that x can assume:
 - (a) Let x = the sum of the three dice.
 - (b) Let x = the minimum number of dots showing on any die.
 - (c) Let x = two times the number of dots on the green die.
- 6.4 One hundred voters in a suburban town are to be selected and asked whether they approve of a civic arena being built. Let x = the number of voters in the sample who approve.
 - (a) What values can x assume?
 - (b) Is x a discrete or continuous random variable?
- 6.5 A manufacturer of light bulbs is interested in various properties about the length of life of his product. Let x = length of life of a light bulb (measured in hours).
 - (a) What values can x assume?
 - (b) Is x a discrete or continuous random variable?
- 6.6 An engineer is testing samples of steel cables to be used in the construction of a suspension bridge.
 - (a) Describe one random variable of possible interest.
 - (b) Is the random variable discrete or continuous?
- 6.7 A health official wanted to investigate the relationship between females with no children and the incidence of breast cancer. Five hundred women, 40–45 years of age, with no children and no prior breast cancer, are selected. Let x = the number of women who develop breast cancer during the following 5 years.
 - (a) What values can x assume?
 - (b) Is x a discrete or continuous random variable?
- 6.8 A known health effect associated with exposure to elevated levels of radon is an increased risk of developing lung cancer. The alpha track detector is a device for measuring the amount of radon in the air. The amount is measured in picocuries of radon per liter of air (pCi/l). The Environmental Protection Agency (EPA) action level for radon in air is 4.0 pCi/l. An alpha track detector is placed in the cellars of 50 randomly selected homes. Classify the random variable in parts (a) and (b) as continuous or discrete, and state the values it can assume.
 - (a) Let x = the mean pCi/l level of the 50 measurements.
 - (b) Let x = the number of homes that exceed the EPA action level for radon in air.

6.3

DISCRETE PROBABILITY DISTRIBUTIONS

In a study of families with one child, a researcher coded families as follows:

$$x = \begin{cases} 0, & \text{if child is a boy} \\ 1, & \text{if child is a girl} \end{cases}$$

Imagine the experiment of randomly selecting a family with one child and recording whether the child is a boy (B) or a girl (G). The sample space is

$$S = \{B, G\}$$

and x is a random variable on this sample space. We can view x as the number of girls in a randomly selected family with one child (0 or 1). We assume that a boy and a girl are equally likely. Hence, the probability that $x = 0$ (a boy) is $\frac{1}{2}$ and the probability that $x = 1$ (a girl) is $\frac{1}{2}$. We sometimes write

$$P(0) = \frac{1}{2} \quad P(1) = \frac{1}{2}$$

The specification of the probabilities associated with the distinct values of this random variable is called its **probability distribution**. We generalize this idea with the following definition.

Definition The specification of the probabilities associated with the various distinct values of a discrete random variable is called a *discrete probability distribution*. The probability associated with the value x is denoted by the symbol $P(x)$.

If many one-child families were selected, and the number of girls (0 or 1) were recorded for each family, a population of data values would be generated. The probability distribution for x (the number of girls) provides a theoretical description of what this population should look like: Since the probability of observing a 0 is $\frac{1}{2}$ and the probability of observing a 1 is $\frac{1}{2}$, the 0's and 1's should occur in equal proportions (in the long run).

It is often convenient to think of a conceptual population of data values associated with a random variable x . This is the collection of values x that would result if the experiment were performed many times. Sometimes we imagine the experiment as being performed repeatedly for an unlimited number of times. In this case, the population is viewed as infinite, because there is no reason to impose a specific limit on the number of times the experiment is performed.

EXAMPLE 6.3

In Chapter 5, we discussed the experiment where a student randomly guesses at the two questions on a true-false quiz. Let x = the number of correct guesses. Find the probability distribution for x .

Solution

Recall that correct on the first question and wrong on the second is CW. The sample space associated with the experiment of taking the quiz by guessing is

$$S = \{CW, WC, WW, CC\}$$

Each outcome in S is equally likely. Therefore, each has the probability $\frac{1}{4}$. The values of the random variable x are given by the following table:

Outcome	x
CW	1
WC	1
WW	0
CC	2

The value 0 will occur if WW occurs. Thus

$$P(0) = P(WW) = \frac{1}{4}$$

The value 1 will occur if either CW or WC occurs—that is, if the event $\{CW, WC\}$ occurs:

$$P(1) = P(\{CW, WC\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

The value 2 will occur if CC occurs. Therefore

$$P(2) = P(CC) = \frac{1}{4}$$

The probability distribution is summarized in the following table:

x	$P(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

where x is the number of correct guesses.

EXAMPLE 6.4

A college statistics class has 20 students. The ages of these students are as follows: One student is 16 years old, four are 18, nine are 19, three are 20, two are 21, and one is 30. Let x = the age of any student (randomly selected). Find the probability distribution for x .

Solution

Since each student has an equal likelihood of being selected, the probability of selecting a particular student is $\frac{1}{20}$. The probability of selecting a student that is 19 years old is $P(19) = \frac{9}{20}$, since there are nine students of that age. The probability distribution is summarized in the following table:

x	$P(x)$
16	$\frac{1}{20}$
18	$\frac{4}{20}$
19	$\frac{9}{20}$
20	$\frac{3}{20}$
21	$\frac{2}{20}$
30	$\frac{1}{20}$

In reading the preceding examples, you may have noted the following properties, which generalize to all discrete probability distributions.

1. Since $P(x)$ is a probability, it will always be a number between 0 and 1 inclusive:

$$0 \leq P(x) \leq 1$$

2. The sum of the values of $P(x)$ for each distinct value of x is 1:

$$\sum P(x) = 1$$

Another point worth noting is this: Suppose in Example 6.3 that we were interested in the probability that the number of correct guesses on the two-question quiz is either 0 or 1. We write this probability as $P(x = 0 \text{ or } x = 1)$. Now x will be 0 for two wrong guesses, WW , and x will be 1 with the guesses CW or WC . So $x = 0$ or $x = 1$ corresponds to the event $\{WW, CW, WC\}$. Notice that

$$\begin{aligned} P(x = 0 \text{ or } x = 1) &= P(\{WW, CW, WC\}) = \frac{3}{4} = \frac{1}{4} + \frac{1}{2} \\ &= P(0) + P(1) \end{aligned}$$

This result generalizes to any discrete distribution:

$$P(x = a \text{ or } x = b) = P(a) + P(b)$$

Every discrete probability distribution must satisfy properties 1 and 2 in the preceding box. Conversely, if we are given values for x and $P(x)$ such that properties 1 and 2 are satisfied, then these values define a discrete probability distribution for some random variable x . The following example illustrates this point.

EXAMPLE 6.5

Consider the expression $P(x)$ defined by the equation

$$P(x) = \frac{x}{6} \quad \text{for } x = 1, 2, 3$$

- (a) Show that $P(x)$ defines a probability distribution:
- (b) Give an example of an experiment, a sample space, and a random variable on the sample space for which $P(x)$ is the probability distribution.

Solution

- (a) To show that $P(x)$ is a probability distribution, we need only show that the two basic properties of a distribution are satisfied.

(1) Clearly, $0 \leq \frac{x}{6} \leq 1$ for $x = 1, 2, 3$.

$$(2) \quad \begin{aligned} \sum P(x) &= P(1) + P(2) + P(3) \\ &= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} = 1 \end{aligned}$$

Therefore, $P(x)$ is a probability distribution.

- (b) Construct an experiment and sample space as follows: Place six balls in a box. Label one ball with the number 1, label two balls with the number 2, and three balls with the number 3. Let the experiment be to select a ball at random. The sample space consists of the six balls. For the random variable, we would let x be the number on the ball selected. It should be clear that $P(x) = x/6$ gives the probability associated with each value of x .

Graphic Representations

We can represent a probability distribution graphically by constructing a type of bar graph called a **probability histogram**. This is constructed by displaying the possible distinct values of the random variable along a horizontal axis. Above each value x of the random variable, we draw a vertical bar having height equal to the probability $P(x)$. We will restrict our attention to those cases where the possible values of x are whole numbers. In such cases, we make the width of each bar 1.

EXAMPLE 6.6

Construct a probability histogram for the distribution of Example 6.3.

Solution

The probability distribution is given by

x	$P(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

The probability histogram is shown in Figure 6.2.

Remark In Example 6.6, notice that the probability associated with a value x is not

only the height of the bar above x , but it is also equal to the area of the bar above x , since the width of each bar is 1. That is,

$$P(x) = \text{area of the bar above } x$$

This will be the case for all discrete distributions we consider. This seemingly insignificant observation will be of great importance in future applications.

EXERCISES

- 6.9 An electronics firm employs 15 people whose yearly salaries (in thousands of dollars) are

25.0 25.7 25.7 26.6 27.2 27.3 27.3 27.3
27.3 27.8 27.8 28.4 28.4 29.0 29.0

Let x be the yearly salary for any employee. Find the probability distribution for x .

- 6.10 Ten thousand Instant Money lottery tickets were sold. One ticket has a face value of \$1000, five tickets have face values of \$500 each, 20 tickets are worth \$100 each, 500 are worth \$1 each, and the rest are losers. Let x = the value of a ticket that you buy. Find the probability distribution for x .
- 6.11 The following table gives information on the range of holdings for 50 major U.S. college and university libraries in 1991–1992 (*Source: The 1994 Information Please Almanac*, 1994, p. 864). The value of x gives a code for each range, with 1 representing the lowest range, and 9 representing the highest range. Find the probability distribution of x .

Number of Libraries	Range of Holdings	x
20	2–3 million	1
12	3–4 million	2
4	4–5 million	3
5	5–6 million	4
5	6–7 million	5
1	7–8 million	6
1	8–9 million	7
1	9–10 million	8
1	10 million or more	9

Note: 2–3 million means at least 2 million and less than 3 million.

- 6.12 A box has five tickets numbered 1, 2, 3, 4, and 5. Two are to be randomly selected without replacement. Let x be the number of occurrences of either 4 or 5. For example, if 2 and 5 are selected, then $x = 1$. If both 4 and 5 are selected, then $x = 2$. Find the probability distribution for x .
- 6.13 The following data are characteristics of the voting-age population regarding the 1992 presidential election in the United States. Number of persons is measured in thousands.

	Number Voted	Number Did Not Vote
Males	53,312	35,245
Females	60,554	36,573

Find the probability distribution of x in parts (a) and (b):

- (a) A name is to be randomly selected from those who voted. Let $x = 0$ if a male is selected. Otherwise, let $x = 1$.
- (b) A name is to be randomly selected from the female voting-age population. Let $x = 0$ if the selected female voted. Otherwise, let $x = 1$.

- 6.14 For adults without known heart disease, a desirable total cholesterol level is less than 200. Between 200 and 239 is categorized as borderline, and above 239 is too high. The following data are from a sample of 825 participants in the Framingham Heart Study with no prior heart disease:

Category	Frequency
Desirable	234
Borderline	297
Too high	294

Let x equal 1, 2, or 3 for desirable, borderline, or too high, respectively. Find the probability distribution for x .

- 6.15 In a population of 3000 people, 1470 were classified as blood type O, 1140 as type A, 300 as type B, and the remaining as type AB. Let $x = 0, 1, 2$, or 3 if a person's blood type is O, A, B, or AB, respectively. Find the probability distribution for x .

In Exercises 6.16–6.19, determine whether each is a probability distribution. Give reasons for your answer.

6.16

x	$P(x)$
2	$\frac{1}{2}$
4	$\frac{1}{4}$
6	$\frac{1}{2}$

6.17

x	$P(x)$
3	$\frac{2}{5}$
9	$\frac{2}{5}$
11	$\frac{2}{5}$
13	$-\frac{1}{5}$

6.18

x	$P(x)$
-3	$\frac{1}{3}$
-1	$\frac{1}{3}$
5	$\frac{1}{3}$

6.19

x	$P(x)$
0	$\frac{2}{3}$
1	$\frac{1}{6}$
2	$\frac{1}{6}$

- 6.20 The probability distribution for the number of electronic instruments produced per hour by an assembly line is as follows:

x	$P(x)$
1	$\frac{1}{36}$
2	$\frac{3}{36}$
3	$\frac{5}{36}$
4	$\frac{7}{36}$
5	$\frac{9}{36}$
6	$\frac{11}{36}$

Construct a probability histogram.

- 6.21** A business executive was sensitive to the possibility of being accused of religious discrimination in his hiring practices. He found that of 600 employees, 204 were Catholic, 228 were Protestant, 72 were Jewish, with the rest classified as Other. Let $x = 0, 1, 2$, or 3 if a person's religion is Catholic, Protestant, Jewish, or Other, respectively. Find the probability distribution and construct the probability histogram.

6.4

MEAN AND VARIANCE

In Example 6.3, we discussed an experiment that consisted of a student randomly guessing at the answers to a two-question true–false quiz. The random variable was x = number of correct guesses (0, 1, or 2). The probability distribution was given by

x	$P(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

The population for this random variable would be the string of values of x that would occur if this experiment were performed over and over by many students. The probability distribution provides a theoretical description of this population, and we ought to be able to use it to predict the mean and variance of all these values.

We will try to simulate this population by considering what the data values should theoretically look like if the experiment is performed by a large number of students, say $N = 40,000$ students, and the number x of correct guesses is recorded for each student. Examining the probability distribution $P(x)$, we see that in an ideal situation we should expect 0 to occur about $\frac{1}{4}$ of the time. Thus we expect about 10,000 zeros. Similarly, we expect about 20,000 ones and 10,000 twos.

x = Number Correct	Frequency
0	10,000
1	20,000
2	10,000

Therefore, we expect the mean of all these values to be

$$\begin{aligned}
\mu &= \frac{\text{sum of all } x \text{ values}}{N} \\
&= \frac{(0)(10,000) + (1)(20,000) + (2)(10,000)}{40,000} \\
&= (0) \cdot \left(\frac{10,000}{40,000}\right) + (1) \cdot \left(\frac{20,000}{40,000}\right) + (2) \cdot \left(\frac{10,000}{40,000}\right) \\
&= (0) \cdot \left(\frac{1}{4}\right) + (1) \cdot \left(\frac{1}{2}\right) + (2) \cdot \left(\frac{1}{4}\right) = 1
\end{aligned}$$

Notice that

$$(0) \cdot \left(\frac{1}{4}\right) = (0) \cdot P(0) \quad (1) \cdot \left(\frac{1}{2}\right) = (1) \cdot P(1) \quad (2) \cdot \left(\frac{1}{4}\right) = (2) \cdot P(2)$$

So we can write

$$\mu = 0 \cdot P(0) + 1 \cdot P(1) + 2 \cdot P(2) = 1$$

This is the sum of the values of $x \cdot P(x)$ for each distinct value of x . We write this as

$$\mu = \sum x \cdot P(x)$$

(which works out to be 1 for the random variable under discussion).

The equation for the mean does not involve the population size N anywhere. In fact, no matter how large the value of N , we would be led by the preceding reasoning to the same formula for predicting the mean. Therefore, we will agree that the mean will be given by this formula.

We generalize this idea in the following definition.

Definition Given a discrete random variable x with probability distribution $P(x)$, the *mean* of the random variable x is defined as

$$\mu = \sum x \cdot P(x)$$

Note: μ is also referred to as the mean of the probability distribution $P(x)$.

We could go through a discussion very similar to the one just given to discover a suitable formula for population variance in terms of $P(x)$. However, this would be unnecessarily repetitious. Perhaps the simplest thing to do is to observe that whereas the population mean is an average of x values, the population variance is an average of values of $(x - \mu)^2$. Then the same discussion that led to a formula for the mean, μ , would lead to a similar formula for σ^2 . The only difference is that instead of summing the values of $x \cdot P(x)$, we would sum values of $(x - \mu)^2 \cdot P(x)$. This leads to the following definitions.

Definitions Given a discrete random variable x with probability distribution $P(x)$, the *variance* of the random variable x is defined as

$$\sigma^2 = \sum (x - \mu)^2 \cdot P(x)$$

The *standard deviation* σ is the square root of the variance. We also use the phrase variance or standard deviation of the distribution $P(x)$.

We often do not know enough about the population to find $P(x)$. In this case, we might obtain a sample from the population and use the sample mean \bar{x} to estimate μ , and s^2 to estimate σ^2 .

EXAMPLE 6.7

Consider the experiment of randomly selecting a family with one child. Let x = the number of girls (0 or 1). For this random variable, calculate the mean, variance, and standard deviation.

Solution

Assuming a girl and boy are equally likely, the probability distribution is given by

x	$P(x)$
0	$\frac{1}{2}$
1	$\frac{1}{2}$

Now

$$\begin{aligned}\mu &= \sum x \cdot P(x) = 0 \cdot P(0) + 1 \cdot P(1) \\ &= (0) \cdot \left(\frac{1}{2}\right) + (1) \cdot \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)\end{aligned}$$

By definition, the variance is

$$\begin{aligned}\sigma^2 &= \sum (x - \mu)^2 \cdot P(x) \\ &= (0 - \frac{1}{2})^2 \cdot P(0) + (1 - \frac{1}{2})^2 \cdot P(1) \\ &= (\frac{1}{4}) \cdot (\frac{1}{2}) + (\frac{1}{4}) \cdot (\frac{1}{2}) = \frac{1}{4} \\ \sigma &= \sqrt{\frac{1}{4}} = \frac{1}{2}\end{aligned}$$

There is an **alternative formula for σ^2** that is a little simpler computationally, but which gives the same result. The formula is

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

This formula tells us to square each distinct value of x , multiply this by $P(x)$, and then add all these products. We then subtract μ^2 from this sum. *In future work, we recommend using the alternative formula unless μ is a whole number.*

EXAMPLE 6.8

For the random variable of Example 6.7 (number of girls in a one-child family), calculate the variance and standard deviation using the alternative formula.

Solution

We saw in Example 6.7 that $\mu = \frac{1}{2}$. Hence

$$\begin{aligned}\sigma^2 &= \left[\sum x^2 \cdot P(x) \right] - \mu^2 = 0^2 \cdot P(0) + 1^2 \cdot P(1) - \left(\frac{1}{2}\right)^2 \\ &= (0) \cdot \left(\frac{1}{2}\right) + (1) \cdot \left(\frac{1}{2}\right) - \left(\frac{1}{4}\right) = \frac{1}{4} \\ \sigma &= \sqrt{\frac{1}{4}} = \frac{1}{2}\end{aligned}$$

Sometimes it is helpful to organize our computations in the form of a table. This is especially recommended if there are more than, say, four or five values for x . We illustrate how such a table is used in the following example.

EXAMPLE 6.9

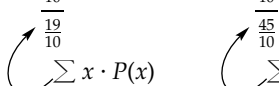
Consider the following probability distribution:

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

We will use Table 6.1 to find the mean, variance, and standard deviation.

Table 6.1

x	$P(x)$	$x \cdot P(x)$	$x^2 \cdot P(x) = x[x \cdot P(x)]$
1	$\frac{4}{10}$	$\frac{4}{10}$	$\frac{4}{10}$
2	$\frac{4}{10}$	$\frac{8}{10}$	$\frac{16}{10}$
3	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{9}{10}$
4	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{16}{10}$
Totals		$\frac{19}{10}$	$\frac{45}{10}$


 $\sum x \cdot P(x)$ $\sum x^2 \cdot P(x)$

$$\begin{aligned}\mu &= \sum x \cdot P(x) = \frac{19}{10} = 1.9 \\ \sigma^2 &= [\sum x^2 \cdot P(x)] - \mu^2 = \frac{45}{10} - (1.9)^2 = 4.5 - 3.61 = .89 \\ \sigma &= \sqrt{.89} \doteq .94\end{aligned}$$

Remark Note that the mean of the distribution in Example 6.9 is 1.9. This is different from the mean of the individual x values (1, 2, 3, 4), which is 2.5. Occasionally, the mean of the distribution is the same as the mean of the individual x values. One example is when the distribution is symmetric. (See the distribution discussed at the beginning of this section.)

Expected Value

Expected value is just another name for the mean of a random variable. Using this name for the mean is often helpful, because many problems are stated in terms of expectation when the mean is involved. For example, an insurance company might be interested in its expected earnings per customer. This is really just the average.

Definition For a random variable x , we define the *expected value* of the random variable, denoted by $E(x)$, to be the mean of the random variable. Therefore, for a discrete random variable,

$$E(x) = \mu = \sum x \cdot P(x)$$

EXAMPLE 6.10

An insurance company sells a life insurance policy with a face value of \$1000 and a yearly premium of \$20. If .1% of the policyholders can be expected to die in the course of a year, what would be the company's expected earnings per policyholder in any year?

Solution

Let x = the amount of money earned by the company from an arbitrary (randomly selected) policyholder in a year. If the policyholder survives the year, then $x = \$20$. If the policyholder dies, the company must pay out \$1000. This minus the \$20 premium means that the company loses \$980. In other words, on this policyholder the company earns $-\$980$ (i.e., $x = -\$980$). The probability that the policyholder dies (i.e., $x = -\$980$) is .001. So the probability that the policyholder lives (i.e., $x = 20$) is .999. Therefore, the probability distribution is

x	$P(x)$
20	.999
-980	.001

The expected earnings per policyholder are

$$\begin{aligned}
 E(x) &= \sum x \cdot P(x) \\
 &= (20)(.999) + (-980)(.001) \\
 &= 19.98 - .98 = 19
 \end{aligned}$$

The company can expect to earn \$19 per policyholder (on the average).

EXAMPLE 6.11 (THE NUMBERS GAME*)

This game consists of selecting a three-digit number. If you guess the right number, you are paid \$700 for each dollar you bet. Each day there is a new winning number. If a person bets \$1 each day for 1 year, how much money can he expect to win (or lose)?

Solution

Let x = the amount won on a given day. If the correct number is guessed, then he wins \$700 – \$1 = \$699. (Do not forget that it costs \$1 to bet.) Otherwise, he loses, so his winnings in this case would be –\$1. There are 1000 possible three-digit numbers: 000 to 999. The probability of winning (selecting the correct three digits) is $\frac{1}{1000} = .001$. The probability of losing is $\frac{999}{1000} = .999$. The probability distribution is

x	$P(x)$
–1	.999
699	.001

Now we find the *expected daily winnings*, $E(x)$:

$$\begin{aligned}
 E(x) &= \sum x \cdot P(x) = (-1)(.999) + (699)(.001) \\
 &= -.999 + .699 = -\$.30
 \end{aligned}$$

Thus he can expect to lose 30 cents per day (on the average). For the year, he can expect to lose $(.30)(365) = \$109.50$.

EXAMPLE 6.12†

A clinic tests blood for a disease that occurs in 1% of the population. The blood samples arrive in batches of 50. The clinic director wondered whether a portion of the blood from each vial should be taken and one test performed on the pooled samples. Then if the test were negative, all samples would be negative. If the test were positive, then each sample would be tested to see which were positive. Under

*For a discussion of this and other casino games, see Koshy, T., *Finite Mathematics and Calculus with Applications*, Santa Monica, Calif.: Goodyear, 1979, Section 4.2.

†This application was given by John Allen Paulos in *Innumeracy—Mathematical Illiteracy and Its Consequences*, New York: Vintage Books, 1990, p. 47. The same problem is discussed for various sample sizes by Richard J. Larsen and Morris L. Marx in *Introduction to Mathematical Statistics and Its Applications*, Englewood Cliffs, N.J.: Prentice-Hall, 1986, p. 158.

this procedure, what is the expected number of tests that will be performed?

Solution

Let x = number of tests. If all 50 people are free of the disease, only one test will be performed ($x = 1$). If the pooled samples test positive, another 50 tests will be done, so $x = 51$. So

$$x = \begin{cases} 1, & \text{if all 50 people are negative} \\ 51, & \text{if at least one person is positive} \end{cases}$$

Since 1% of the population has the disease, 99% do not. So the probability that one person will be negative (does not have the disease) is .99. The probability that all 50 people are negative is the probability that

(first is negative) *and* (second is negative) *and* ... *and* (50th is negative)

Using the Multiplication Rule for Independent Events, the probability of this is $(.99)^{50}$. At least one person being positive is the opposite (or complement) of all being negative. Hence the probability that at least one person is positive is $1 - (.99)^{50}$. The probability distribution for x is

x	$P(x)$
1	$(.99)^{50}$
51	$1 - (.99)^{50}$

The expected number of tests is

$$\begin{aligned} E(x) &= \sum x \cdot P(x) = (1)(.99)^{50} + (51)[1 - (.99)^{50}] \\ &= .605006 + 20.144691 = 20.749697 \end{aligned}$$

So by pooling the samples, we expect only about 21 tests to be performed on average as opposed to 50 tests if each sample were to be tested.

EXERCISES

For the probability distributions in Exercises 6.22–6.25, find (a) the mean, (b) the variance, and (c) the standard deviation. Also, (d) describe each of the distributions as skewed to the left, skewed to the right, or symmetric. (To organize your computations, it may help to use a table like Table 6.1.)

6.22

x	$P(x)$
0	$\frac{1}{10}$
1	$\frac{2}{10}$
2	$\frac{3}{10}$
3	$\frac{4}{10}$

6.23

x	$P(x)$
0	$\frac{4}{10}$
1	$\frac{3}{10}$
2	$\frac{2}{10}$
3	$\frac{1}{10}$

6.24

x	$P(x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

6.25

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

6.26 A dentist has determined that the number of patients x treated in an hour is described

by the probability distribution given here. Find (a) the mean, (b) the variance, and (c) the standard deviation.

x	$P(x)$
1	$\frac{2}{15}$
2	$\frac{10}{15}$
3	$\frac{2}{15}$
4	$\frac{1}{15}$

- 6.27 The manager of a baseball team has determined that the number of walks x issued in a game by one of the pitchers is described by the probability distribution given here. Find (a) the mean, (b) the variance, and (c) the standard deviation.

x	$P(x)$
0	$\frac{1}{20}$
1	$\frac{2}{20}$
2	$\frac{3}{20}$
3	$\frac{11}{20}$
4	$\frac{3}{20}$

- 6.28 Appendix Table A describes how to select random numbers from Appendix Table B.1, a table of random numbers. Let x = a randomly selected single digit.

(a) Assume that the probability distribution of x is as follows:

$$P(x) = \frac{1}{10} \quad \text{for } x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$$

- (i) Find the mean of x .
 (ii) Find the standard deviation of x .
- (b) Randomly select 25 digits from Appendix Table B.1. Calculate the sample mean and the sample standard deviation. Compare these sample results with the population mean and standard deviation in part (a).
- 6.29 An altered die has one dot on one face, two dots on three faces, and three dots on two faces. The die is to be tossed once. Let x be the number of dots on the upturned face. Find the mean and variance of x .
- 6.30 A card is to be selected from an ordinary deck of 52 cards. Suppose that a casino will pay \$10 if you select an ace. If you fail to select an ace, you are required to pay the casino \$1.
- (a) If you play this game once, how much money does the casino expect to win?
 (b) If you play the game 26 times, how much money does the casino expect to win?
- 6.31 In the game of craps, a player rolls two dice. If the first roll results in a sum of 7 or 11, the player wins. If the first roll results in a 2, 3, or 12, the player loses. If the sum on the first roll is 4, 5, 6, 8, 9, or 10, the player keeps rolling until he throws a 7 or

the original value. If the outcome is a 7, the player loses. If it is the original value, the player wins. The probability that a player will win is .493. Suppose that a player pays \$5 to a casino if he loses and is paid \$4 for a win. What is the expected loss for the player if he plays (a) one game? (b) ten games?

- 6.32 A bus company is interested in two potential contracts, one for an express and the other for local stops. The probabilities that the bids will be accepted are .70 and .50 with costs of \$500 and \$750, respectively. The estimated total incomes are \$6000 and \$10,000, respectively. If the company were allowed only one bid, which bid should they enter?
- 6.33 A high school class decides to raise some money by conducting a raffle. The students plan to sell 2000 tickets at \$1 apiece. They will give one prize of \$100, two prizes of \$50, and three prizes of \$25. If you plan to purchase one ticket, what are your expected net winnings? (*Hint:* The probability of getting the \$100 ticket is $\frac{1}{2000}$, of getting a \$50 ticket is $\frac{2}{2000}$, and of getting a \$25 ticket is $\frac{3}{2000}$.)
- 6.34 In Exercises 5.16 and 5.56, we considered Megabucks, a lottery game conducted by the Massachusetts State Lottery Commission. Megabucks consists of selecting six numbers from the 42 numbers $\{1, 2, 3, 4, \dots, 41, 42\}$ (no repetitions and order does not matter). The commission selects by chance the six winning numbers. You pay \$1 to play. You get a free ticket if three of your numbers match three of the winning numbers, \$75 for matching four numbers, \$1500 for matching five numbers. Assume you get \$1,000,000 if you match all six numbers. The sample space S consists of all possible six-number combinations that can be selected. It can be shown that $n(S) = 5,245,786$. The number of elements in S that match 0, 1, 2, 3, 4, 5, or 6 winning numbers is given here. What are your expected winnings?

Number of Matches	Number of Possibilities
0	1,947,792
1	2,261,952
2	883,575
3	142,800
4	9,450
5	216
6	1

6.5

THE BINOMIAL PROBABILITY DISTRIBUTION

Consider the experiment of rolling a single die and recording whether the number (of dots) on the top face is odd (O) or even (E). This experiment has two possible outcomes: O or E . We can construct a new experiment by performing this basic experiment over and over a certain number of times. (In a situation like this, the basic experiment is called a *trial*.) For example, suppose that we perform this basic experiment three times. If we observed an even number, then an odd, and then an even, we would represent this outcome by EOE . The sample space for the new experiment is

$$\{OOO, EOO, OEO, OOE, OEE, EOE, EEO, EEE\}$$

This experiment consisted of repeating a trial a certain number of times; each trial had two possible outcomes. Such an experiment is called a **binomial experiment**.

There are many practical situations that are in essence binomial experiments. For example, we may interview 10,000 voters to see how many favor candidate A.

Here the trial is to interview a voter. This trial has two possible outcomes: Voter favors candidate A or does not favor candidate A. The trial is repeated 10,000 times. As another example, a manufacturer of transistors may select 1000 transistors, then repeatedly perform the trial of checking each transistor to see whether it is defective.

We summarize these ideas in the following definition.

Definition A *binomial experiment* is an experiment that has the following properties:

1. It consists of performing some basic experiment a fixed number of times, n . Each time the basic experiment is performed, we call this a *trial*.
2. Each trial is identical and has two possible outcomes. We arbitrarily call one outcome success (S) and the other failure (F). We call the probability of success $P(S) = p$ and the probability of failure $P(F) = q$. Note that $p + q = 1$. (So $q = 1 - p$.) The values of p and q do not change from one trial to another.
3. The trials are independent of one another, that is, the outcome of one trial has no bearing on the outcome of any other trial.

The random variable of interest, the *binomial random variable* x , is the number of successes in n trials. Note that x can assume the values $0, 1, 2, \dots, n$. The probability distribution of x is called a *binomial probability distribution*.

EXAMPLE 6.13

Consider the binomial experiment of rolling a die three times. Each time we record whether the number of dots showing is odd (O) or even (E). Let x = the total number of evens recorded. Find the binomial probability distribution $P(x)$.

Solution

The trial is to roll the die and record O or E . We will call E a success. So $S = E$ and failure $F = O$. Since half the numbers on a die are even and half are odd, $p = P(S) = \frac{1}{2}$ and $q = P(F) = \frac{1}{2}$. We have seen that the sample space is

$$\{OOO, EOO, OEO, OOE, OEE, EOE, EEO, EEE\}$$

Each outcome is equally likely and therefore has the probability $\frac{1}{8}$. The binomial random variable is x = the number of successes = the number of evens. We must have $x = 0, 1, 2$, or 3 .

Now $x = 0$ corresponds to OOO , so $P(0) = P(OOO) = \frac{1}{8}$. The value $x = 1$ corresponds to the event $\{EOO, OEO, OOE\}$. So

$$P(1) = P(\{EOO, OEO, OOE\}) = \frac{3}{8}$$

The value $x = 2$ corresponds to the event $\{OEE, EOE, EEO\}$. Therefore,

$$P(2) = P(\{OEE, EOE, EEO\}) = \frac{3}{8}$$

Finally, the value $x = 3$ corresponds to EEE , so that $P(3) = P(EEE) = \frac{1}{8}$. The binomial probability distribution for this situation is summarized as follows:

x	$P(x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

The essential mathematical ingredients of a binomial experiment are the number of trials n , the probability of success on an individual trial p , and the number of successes x . It would be convenient if we had a formula for the binomial distribution, that is, a formula for $P(x)$, using these mathematical ingredients. Such a formula does exist.

Given a binomial experiment consisting of n trials where the probability of success on an individual trial is p and the probability of failure is q (where $q = 1 - p$), it can be shown that the probability of exactly x successes in n trials is given by the following formula*:

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

The symbol $n!$ (read “ n factorial”) is defined for any positive integer to be

*If you have read Sections 5.4 and 5.6, it is not difficult to see why this formula works. First, consider the event consisting of success on the first x trials and failure on the next $n - x$ trials:

$$\underbrace{SS \cdots SS}_{x} \underbrace{FF \cdots FF}_{n-x} \leftarrow n \text{ trials}$$

The trials are independent. Hence the Multiplication Rule for independent events tells us the probability of this is

$$\underbrace{p \cdot p \cdots p}_{x} \cdot \underbrace{q \cdot q \cdots q}_{n-x} = p^x \cdot q^{n-x}$$

Now there are a number of different arrangements that can result in x successes and $n - x$ failures, all having the same probability, $p^x q^{n-x}$. To count how many ways this can happen, just calculate how many ways we can specify which x trials out of the n trials are the successes. In Section 5.6, we saw that this number is

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

So there are this many ways of getting x successes and $n - x$ failures, all having probability $p^x q^{n-x}$. Hence, the probability of x successes in n trials is

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

$$n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$$

Therefore

$$3! = 1 \cdot 2 \cdot 3 = 6$$

$$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$$

$$1! = 1$$

We will agree that $0! = 1$.

The values of $P(x)$ constitute the **binomial probability distribution**. The word *binomial* is used because the values of $P(x)$ for the various values of x are precisely the terms in the binomial expansion of $(p + q)^n$. For example, when $n = 2$, the values of $P(x)$ for $x = 0, 1$, and 2 are the three terms on the right-hand side of the equation

$$(p + q)^2 = p^2 + 2pq + q^2$$

In Figure 6.3, we display probability histograms for two binomial probability distributions.

EXAMPLE 6.14

Verify that the formula for $P(x)$ in the preceding box does give the probability distribution for the binomial variable of Example 6.13.

Solution

For this experiment, $n = 3$. Also, success = record even (E) and failure = record odd (O). Thus $p = P(E) = \frac{1}{2}$ and $q = P(O) = 1 - p = \frac{1}{2}$. Finally, x = number of evens recorded in three rolls of the die. Therefore,

$$P(x) = \frac{3!}{x!(3-x)!} \cdot \left(\frac{1}{2}\right)^x \cdot \left(\frac{1}{2}\right)^{3-x}$$

Observe that

$$\left(\frac{1}{2}\right)^x \cdot \left(\frac{1}{2}\right)^{3-x} = \left(\frac{1}{2}\right)^{x+3-x} = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

and so

$$P(x) = \frac{3!}{x!(3-x)!} \cdot \frac{1}{8}$$

Now

$$P(0) = \frac{3!}{0!(3-0)!} \cdot \frac{1}{8} = \frac{6}{(1) \cdot (6)} \cdot \frac{1}{8} = \frac{1}{8}$$

$$P(1) = \frac{3!}{1!(3-1)!} \cdot \frac{1}{8} = \frac{6}{(1) \cdot (2)} \cdot \frac{1}{8} = \frac{3}{8}$$

$$P(2) = \frac{3!}{2!(3-2)!} \cdot \frac{1}{8} = \frac{6}{(2) \cdot (1)} \cdot \frac{1}{8} = \frac{3}{8}$$

$$P(3) = \frac{3!}{3!(3-3)!} \cdot \frac{1}{8} = \frac{6}{(6) \cdot (1)} \cdot \frac{1}{8} = \frac{1}{8}$$

These results agree with the results of Example 6.13.

EXAMPLE 6.15

Assume that when a certain hunter shoots at a pheasant, the probability of hitting it is .6. Find the probability that the hunter

- (a) Will hit exactly four of the next five pheasants at which he shoots.
- (b) Will hit at least four of the next five.
- (c) Will hit at least one of the next five.

Solution

- (a) trial = shoots at a pheasant
 success = hits the pheasant
 failure = does not hit the pheasant
 $P(S) = p = .6$
 $P(F) = q = 1 - p = .4$

The trial is repeated five times, so $n = 5$. The binomial distribution giving the probability of exactly x successes in five trials is

$$P(x) = \frac{5!}{x!(5-x)!} \cdot (.6)^x \cdot (.4)^{5-x}$$

The probability of hitting exactly four of the next five pheasants is

$$\begin{aligned} P(4) &= \frac{5!}{4!(5-4)!} \cdot (.6)^4 \cdot (.4)^{5-4} \\ &= \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{(1 \cdot 2 \cdot 3 \cdot 4) \cdot (1)} \cdot (.6)^4 \cdot (.4) = (5) \cdot (.6)^4 \cdot (.4) \\ &= .25920 \end{aligned}$$

- (b) Now the probability that he hits at least four of the next five pheasants is $P(x=4 \text{ or } x=5)$. But we know that

$$\begin{aligned} P(x=4 \text{ or } x=5) &= P(4) + P(5) \\ &= .25920 + P(5) \end{aligned}$$

and

$$\begin{aligned} P(5) &= \frac{5!}{5!(5-5)!} \cdot (.6)^5 \cdot (.4)^{5-5} \\ &= \frac{5!}{5!0!} \cdot (.6)^5 \cdot (.4)^0 \end{aligned}$$

Keeping in mind that $0! = 1$ and $(.4)^0 = 1$, we get

$$P(5) = (.6)^5 = .07776$$

Thus

$$\begin{aligned} P(x=4 \text{ or } x=5) &= P(4) + P(5) \\ &= .25920 + .07776 = .33696 \end{aligned}$$

- (c) Now we find the probability that he hits at least one out of the next five pheasants. Call A the event consisting of *no* hits in the next five shots.

Then the complement of this event, \bar{A} , consists of hitting at least one. Since $P(A) + P(\bar{A}) = 1$, we have

$$P(\text{hitting at least 1}) = P(\bar{A}) = 1 - P(A) = 1 - P(0)$$

Now we need only evaluate $P(0)$ and subtract from 1:

$$\begin{aligned} P(0) &= \frac{5!}{0!(5-0)!} \cdot (.6)^0 \cdot (.4)^{5-0} \\ &= \frac{5!}{5!} \cdot (1) \cdot (.4)^5 = (.4)^5 \\ &= .01024 \end{aligned}$$

Therefore,

$$\begin{aligned} P(\text{hitting at least one}) &= 1 - P(0) \\ &= 1 - .01024 \\ &= .98976 \end{aligned}$$

The formula for calculating binomial probabilities $P(x)$ can be rather unpleasant to use for all but very small values of n . Appendix Table B.2 gives values of $P(x)$ for various values of n and x up to $n = 25$. [In Chapter 7, we develop a method of finding $P(x)$ for values of $n > 25$.]

Let's use Appendix Table B.2 to find $P(x)$ when $n = 5$, $p = .60$, and $x = 4$. In Example 6.15, we found this value to be .259 rounded to three decimal places. In the far left column, locate the desired value of n , in this case $n = 5$. To the right and below this, locate the row containing the desired value of x , here $x = 4$. At the top of the table, locate the column containing the desired value of p , here $p = .60$. Now where the row containing $x = 4$ and the column containing $p = .60$ intersect, we find the value of $P(4)$, namely, .259. (See Table 6.2, which shows a portion of Appendix Table B.2.)

Notice that some of the probabilities in Appendix Table B.2 are represented as 0+. This means that the probability is so small as to be negligible and hence may be assigned the value 0 in calculations.

EXAMPLE 6.16

Thirty percent of the voters in a large voting district are veterans. If 10 voters are randomly selected, find the probability that less than five will be veterans.

Solution

This is a binomial experiment.

trial = select voter

success = voter is a veteran

$$\begin{aligned} P(S) &= p = \text{proportion of voters in district who are veterans} \\ &= .30 \end{aligned}$$

$n = 10$ (repeat the trial 10 times)

$x =$ number of veterans

We want to find $P(x < 5)$. Using Appendix Table B.2, we get

$$\begin{aligned}
 P(x < 5) &= P(0) + P(1) + P(2) + P(3) + P(4) \\
 &= .028 + .121 + .233 + .267 + .200 \\
 &= .849
 \end{aligned}$$

Note: Since the voting district is large, we can safely assume that the value of p remains essentially the same from one trial to the next.

EXAMPLE 6.17

A company that manufactures color television sets claims that only 5% of its sets will need to be adjusted by a technician before being sold. If an appliance dealer sells 20 of these sets, how likely is it that more than three of them will need to be adjusted?

Solution

trial = check one of the sets
 success = it needs adjustment
 $P(S) = p = .05$
 $n = 20$ (repeat the trial 20 times)
 x = number out of 20 that need adjustment

We want to find $P(x > 3)$. Using Appendix Table B.2, add the values of $P(x)$ for x from 4 to 20:

$$\begin{aligned}
 P(x > 3) &= P(4) + P(5) + \cdots + P(19) + P(20) \\
 &= .013 + .002 + (\text{negligible terms}) \\
 &= .015
 \end{aligned}$$

We conclude that if the manufacturer is right, it is very unlikely that more than three sets will need to be adjusted.

Example 6.17 provides some insight into how the binomial distribution might be used in statistical inference. For example, the manager of an appliance store chain received a large shipment of these TV sets, and she wished to test the manufacturer's claim that only 5% (or perhaps less) will need adjustment. She might randomly select 20 sets from the shipment and determine how many of these sets need adjustment. We can view the population as consisting of the entire shipment of TV sets from the manufacturer. The 20 sets then constitute a sample from the population. Based on the number of TV sets out of the 20 that are in need of adjustment, she will decide whether there is strong evidence to suggest that more than 5% of the sets in the population will be in need of adjustment. She would want the evidence to be strong because if she concludes that more than 5% of the population need adjusting, she will return the entire shipment—a drastic step. What criterion might she use? That is, if x = the number of TV sets out of the 20 needing adjustment, how large should x be for the manager to conclude that the shipment should be returned?

Since the manufacturer claims 5% (or less) will need adjusting, it would not be alarming if we found 5% of the 20, or one TV set, needing adjusting. Perhaps two sets would not even be alarming. (After all, we can expect some fluctuation from one

batch of 20 sets to the next.) But the larger the value of x , the more we are inclined to believe that more than 5% of the population will need adjusting. If the value of x is large enough (relative to 1), the manager would reject the manufacturer's claim (and the shipment as well). By "large enough" we mean so large that it would be highly unlikely that we would observe such a large value of x if the manufacturer's claim were true. Note that we have already seen a range of such unlikely values in Example 6.17. We said that it would be very unlikely (only a 1.5% chance) that we would get a value of x greater than 3 if in fact the manufacturer were correct. Therefore, the manager could use this as her criterion: If more than three of the TV sets do need adjusting, the evidence would strongly suggest that the manufacturer's claim should be rejected.

We will see in later chapters that a binomial experiment is a very important concept in a variety of statistical inference problems ranging from industrial quality control to voter preference studies. The reason for this is that *the proportion of elements in a population possessing a certain characteristic of interest may be viewed as the probability of success in a binomial experiment*. In Example 6.16, we discussed the proportion of voters in a district who were veterans. This proportion was .30. A trial consisted of selecting a voter by chance. A success occurs if the voter selected is a veteran. The binomial experiment consisted of performing the trial 10 times. Since 30% of the voters are veterans, the probability that a voter selected on a single trial will be a veteran is .30. This is the probability of success, so that $P(S) = p = .30$. Sometimes such proportions are unknown, and we can use a binomial variable to study these proportions. We investigate these proportions, known as **population proportions**, in Chapter 9.

Mean and Variance of a Binomial Random Variable

Suppose that we consider the binomial experiment of tossing a fair coin $n = 100$ times. Let success = heads. Clearly, $p = .5$. Let x = the number of heads. When the coin is tossed 100 times, how many heads would you expect to get, that is, what is your best guess? We would guess that half (or 50) would be heads. In other words, the expected value of x is 50. This can be obtained from the formula

$$E(x) = n \cdot p = 100 \cdot \left(\frac{1}{2}\right) = 50$$

This formula, in fact, works in general. That is, it can be proved rigorously, using the formula for μ given previously in this chapter, that **for a binomial experiment** consisting of n trials with the probability of success p , the mean or expected value of x is

$$\mu = E(x) = n \cdot p$$

It can also be proved, using the formula for variance given previously in the chapter, that for a binomial variable

$$\sigma^2 = n \cdot p \cdot q$$

Thus, for the number of heads when a coin is tossed 100 times

$$\sigma^2 = n \cdot p \cdot q = (100)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 25$$

$$\sigma = \sqrt{25} = 5$$

EXERCISES

- 6.35** Which of the following are binomial experiments? For those that are not, indicate which part of the definition of a binomial experiment does not apply.
- (a) Tossing a fair coin 1000 times and counting the number of times a head appears
 - (b) Tossing a fair coin and counting the number of tosses before a head appears
 - (c) Checking five students for drug use from a class of 30 students in which 10% use drugs
 - (d) A state agency randomly selects 20 liquor stores, with replacement, and counts the number of stores involved in price fixing
 - (e) A shipment of 40 appliances contains two defectives. A dealer tests 15 of the appliances and counts the number that fail to meet specifications.
- 6.36** Consider a binomial experiment with $n = 4$, $p = .6$, and $x =$ the number of successes. Use the formula for $P(x)$ to find the probability that
- (a) $x = 0$ (b) $x = 1$ (c) $x = 2$
- 6.37** Consider a binomial experiment with $n = 5$, $p = .7$, and $x =$ the number of successes. Use the formula for $P(x)$ to find the probability that
- (a) $x = 3$ (b) $x = 4$ (c) $x = 5$
- 6.38** Consider a binomial experiment with $n = 11$, $p = .4$, and $x =$ the number of successes. Use Appendix Table B.2 to find the probability that
- (a) x is less than 2.
 - (b) x is greater than 5 and less than 8.
 - (c) x is greater than or equal to 5 and less than or equal to 8.
 - (d) x equals 6.
 - (e) x is greater than 0.
- 6.39** Consider a binomial experiment with $n = 9$, $p = .7$, and $x =$ the number of successes. Use Appendix Table B.2 to find the probability that
- (a) x is less than 3.
 - (b) x equals 3.
 - (c) x is greater than 3 and less than 5.
 - (d) x is greater than or equal to 3 and less than or equal to 5.
 - (e) x is less than 9.
- 6.40** Sixty percent of the voters in a large town are opposed to a proposed development. If 20 voters are selected at random, find the probability that
- (a) 10 are opposed to the proposed development.
 - (b) More than 13 are opposed to the proposed development.
 - (c) Less than 10 are opposed to the proposed development.

- 6.41 Forty percent of the student body at a large university are in favor of a ban on drinking in the dormitories. Suppose 15 students are to be randomly selected. Find the probability that
- (a) Seven favor the ban.
 - (b) Fewer than four favor the ban.
 - (c) More than two favor the ban.
- 6.42 It has been reported that 30% of the population of women who had given birth in the last year and had less than a high school education were in the labor force (*Source: The 1994 Information Please Almanac*, 1994, p. 843). In a random sample of 25 from this population, find the probability that
- (a) Ten are in the labor force.
 - (b) The sample will contain 4, 5, 6, or 7 women in the labor force.
- 6.43 Sixty percent of the Framingham Heart Study participants have a total serum cholesterol level below 238. An HMO statistician is interested in interviewing 15 people. If 15 people are randomly selected, find the probability that
- (a) Eight people will have a total serum cholesterol level below 238.
 - (b) Five people or more will have a total serum cholesterol level below 238.
- 6.44 A retailer decides that he will reject a large shipment of light bulbs if there is more than one defective bulb in a sample of size 10. If the defective rate is .10, what is the probability that the retailer will reject the shipment?
- 6.45 A person has a 5% chance of winning a free ticket in a state lottery. If she plays the game 25 times, what is the probability she will win one or more free tickets?
- 6.46 Five percent of the cans of handballs purchased at a sporting goods store are unacceptable. If 12 cans are purchased, find the probability that
- (a) All of the cans are acceptable.
 - (b) More than two cans are unacceptable.
- Suppose 200 cans were bought. What is the expected number of unacceptable cans?
- 6.47 A screening examination is required of all applicants for a technical writing position. The examination consists of 16 questions. Each question has five choices, consisting of the correct answer and four incorrect answers. A curious applicant wonders about some probabilities if she were to randomly guess at each question.
- (a) What is the probability of getting three correct?
 - (b) What is the probability of getting two or more correct?
 - (c) If 50 applicants took the exam and each guessed randomly at all the questions, what would you guess the mean number of correct answers to be?
- 6.48 Thirty-eight percent of the people have blood type A. In a random sample of 20 people, find the probability that
- (a) One will have type A.
 - (b) Two or three will have type A.
 - (c) One or more will not have type A.
- Suppose 100 samples of size 20 were selected, and the number of people in each sample with type A was recorded.
- (d) What should be the approximate mean number of people with type A?
- 6.49 Ten percent of the people have blood type B. In a random sample of 20 people, find the probability that
- (a) Three will have type B.
 - (b) More than five will have type B.

(c) Fewer than two will have type B.

Suppose 100 samples of size 20 were selected, and the number of people in each sample with type B was recorded.

(d) What should be the approximate mean number of people with type B?

6.50 There are 4,230,000 eligible voters in a state, of whom 147,000 are black. A sample of 100 voters is to be selected. Let x be the number of blacks. Assume a binomial experiment. Find (a) the mean, (b) the variance, and (c) the standard deviation of x .

6.51 Between 1972 and 1974, 15 out of 405 teachers (3.7%) hired in Hazelwood, St. Louis County, Missouri, were black. In St. Louis County *plus* the nearby city of St. Louis, 15.4% of teachers were black. The Equal Employment Opportunity Commission (EEOC) sued Hazelwood for discrimination against blacks and won the case in the court of appeals. Assuming that the 405 teachers hired constitute a random sample from a population of teachers, 15.4% of whom are black, answer the following:

(a) What is the mean (or expected) number of black teachers?

(b) What is the standard deviation?

(c) In the case of *Castenada v. Partida*, 430 U.S. 482 (1977), the “Standard Deviation Rule” was advanced. This states that if the observed value (15 teachers in the Hazelwood case) is more than 2 or 3 standard deviations from the expected value, discrimination may be present. How many standard deviations away from the expected value is the observed value of 15?

(d) On appeal to the Supreme Court, the decision of the court of appeals was vacated. The Supreme Court noted that the relevant job market for teachers might well be St. Louis County alone (which does not include the city of St. Louis), where the percentage of black teachers is 5.7%. Redo parts (a), (b), and (c) using this percentage.

(e) This analysis assumes that the selection of the 405 teachers is a random process (with respect to race). Can the selection of teachers be considered a random process?

For a discussion, see DeGroot, M., S. Fienberg, and J. Kadane, *Statistics and the Law*, New York: Wiley, 1986, pp. 1–48.

6.52 Seven percent of printed circuit boards made by the SKC company are defective. A company official wishes to see whether the percentage of defective boards in a current batch has decreased. A sample of size 50 is to be selected. Let x be the number of defectives. Assume a binomial experiment. If there is no change from the 7% figure, find (a) the mean, (b) the variance, and (c) the standard deviation of x .

6.53 Consider a binomial experiment with $n = 4$. Construct probability histograms when

(a) $p = .3$ and $p = .7$. Comment on the relationship between the two.

(b) $p = .4$ and $p = .6$. Comment on the relationship between the two.

6.54 Consider a binomial experiment with $p = .5$. Construct probability histograms when $n = 3$ and $n = 5$. Comment on the shapes of the distributions.

6.55 Consider a binomial experiment with $n = 3$. Construct probability histograms when $p = .2$ and $p = .4$. You will see that each distribution is skewed to the right. Compare the amount of skewness in the two histograms.

6.6

USING MINITAB (OPTIONAL)

PDF Command

The PDF (probability distribution function) command can be used to find prob-

abilities for a variety of probability distributions. In Example 6.15, we found the binomial probability $P(4)$ when there are 5 trials and the probability of success is .6. We can obtain this probability as follows:

<u>Session Commands</u>	<u>Dialog Box</u>
MTB \downarrow PDF 4; SUBC \downarrow BINOMIAL 5 .6.	Calc \blacktriangleright Probability Distributions \blacktriangleright Binomial Type 5 for Number of trials Type .6 for Probability of success Click Input constant . Type 4 in box Click OK

Output

Probability Density Function

Binomial with n = 5 and p = 0.600000

x	P(X j= x)
4.00	0.2592

We see that $P(4) = .2592$. To store this probability for later use, the main command in the Session window would be

PDF 4, STORE IN K1

In the Dialog box, click **Optional storage**, and type *K1* in the box.

We can find several probabilities at a time. Suppose we want $P(4)$ and $P(5)$. Put the values 4, 5 in column 1. The following program would return the values of $P(4)$ and $P(5)$:

<u>Session Commands</u>	<u>Dialog Box</u>
MTB \downarrow PDF C1; SUBC \downarrow BINOMIAL 5 .6.	Calc \blacktriangleright Probability Distributions \blacktriangleright Binomial Type 5 for Number of trials Type .6 for Probability of success Type <i>C1</i> for Input column Click OK

If we wanted to store $P(4)$ and $P(5)$ in, say, column 2, the Session window command would have been

PDF C1, STORE IN C2;

In the Dialog box, we would have typed C2 in the **Optional storage** box. Column 2 would then contain the probabilities:

C2	
0.25920	$\leftarrow P(4)$
0.07776	$\leftarrow P(5)$

We have seen these probabilities in Example 6.15 (where a hunter shoots at five pheasants). We were interested in the probability of at least four hits:

$$P(x = 4 \text{ or } x = 5) = P(4) + P(5)$$

To obtain this, we would just type in the Session window:

```
SUM C2
```

Minitab would print the value 0.33696. In Example 6.15, we found the probability of at least one hit by computing $1 - P(0)$. To compute this using Minitab, first compute $P(0)$ and store it as K1 using the techniques described previously. Then in the Session window type

```
LET K2 = 1 - K1
PRINT K2
```

Other distributions can be specified by a subcommand or by appropriate choices in the Dialog box. For example, consider the distribution

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

Suppose we put the x values in column 1 and the probabilities in column 2. The following program would return the value of $P(2)$:

Session Commands

```
MTB> PDF 2;
SUBC> DISCRETE C1 C2.
```

Dialog Box

```
Calc ► Probability Distributions ► Discrete
Type C1 for Values in
Type C2 for Probs in
Click Input constant. Type 2 in box
Click OK
```

Output

```
Probability Density Function
```

```
Discrete distribution using values in C1 and probabilities in C2
```

x	P(X = x)
2.00	0.4000

CDF Command

Whereas PDF gives probabilities $P(x=k)$, the CDF (cumulative distribution function) gives **cumulative probabilities** $P(x \leq k)$. For example, to find $P(x \leq 4)$ for the binomial distribution discussed previously,

Session Commands	Dialog Box
MTB \downarrow CDF 4; SUBC \downarrow BINOMIAL 5 .6.	Calc \blacktriangleright Probability Distributions \blacktriangleright Binomial Click Cumulative probability Type 5 for Number of trials Type .6 for Probability of success Click Input constant and Type 4 in box Click OK

Output

Cumulative Distribution Function

Binomial with n = 5 and p = 0.600000

x	P(X \leq x)
4.00	0.9222

We see from this that $P(x \leq 4) = .9222$

INVCDF Command

We have seen that CDF gives probabilities of the form $P(x \leq k)$. For example, we saw that for a binomial distribution with $n = 5$, $p = .6$, CDF 4 gave .9222:

$$P(x \leq 4) = .9222$$

INVCDF .9222 would give the value 4. In summary, CDF answers the question

$$P(x \leq k) = \underline{\quad ? \quad}$$

for a given value of k , whereas INVCDF answers the question

$$P(x \leq \underline{\quad ? \quad}) = p$$

for a given probability p where $0 < p < 1$. Actually, this is a bit of an oversimplification. Sometimes there will not be an x value having the exact cumulative probability specified. In this case, Minitab prints the two x values with the closest cumulative probabilities to the one specified. Even when there is an x value having the cumulative probability specified, Minitab prints that x value and the next lower x value, as shown here:

Session Commands

```
MTB> INVCDF .9222;
SUBC> BINOMIAL 5 .6.
```

Dialog Box

Calc ► Probability Distributions ► Binomial
 Click **Inverse cumulative probability**
 Type 5 for **Number of trials**
 Type .6 for **Probability of success**
 Click **Input constant** and Type .9222 in box
 Click **OK**

Output

Inverse Cumulative Distribution Function

Binomial with n = 5 and p = 0.600000

x	P(X j= x)	x	P(X j= x)
3	0.6630	4	0.9222

We see from this that $P(x \leq 4) = .9222$.

If we think of probability distributions as abstract descriptions of populations, the INVCDF command gives percentiles. For example, if for some distribution

$$P(x \leq 46) = .90$$

then the percentage of values in the population less than or equal to 46 is 90%. So 46 is the 90th percentile.

Reminder PDF, CDF, and INVCDF can all store results instead of printing them. Just remember to store constants as constants and columns as columns. For example,

```
INVCDF .43, STORE IN K1;
```

and

```
INVCDF C1, STORE IN C2;
```

Similar remarks apply to the Dialog box.

PRACTICE QUIZ (Answers follow the Review Exercises.)

For the problems in Questions 1–4, give the approximate values that will be returned, without using Minitab.

- PDF 4;
BINOMIAL N = 14, P = .3.
- SET C1
3, 5
END
PDF C1;
BINOMIAL N = 8, P = .7.
- CDF 2, STORE IN K1;
BINOMIAL N = 5, P = .1.

- ```

 LET K2 = 1 - K1
 PRINT K2
4. READ INTO C1 C2
 0 .1
 1 .2
 2 .3
 3 .4
 END
 INVCDF .6;
 DISCRETE, VALUES IN C1, PROBS IN C2.
5. Write a program to find μ , σ^2 , and σ for the distribution in Question 4. (Note: The
 Minitab command for finding square roots is SQRT; it works just like the MEAN
 command.)

```

### EXERCISES

Suggested exercises for use with Minitab are 6.77, 6.78, 6.79, 6.80, and 6.81.

## 6.7

### SUMMARY

In this chapter, we introduced numbers into the setting of a sample space by means of a **random variable**, which is a rule by which we can assign a number to each outcome. Presumably the number will describe some numerical property of the outcome that will be of interest.

A random variable is **continuous** if it can assume any value on some interval or continuous scale; it is **discrete** if the values it can assume constitute a sequence of isolated or separated points on the number axis.

We also studied the notion of the **probability distribution** for a discrete random variable  $x$ . We said  $P(x)$  is the probability of observing the value  $x$ . We then studied the **mean**  $\mu$  and **variance**  $\sigma^2$  for a discrete random variable:

$$\mu = \sum x \cdot P(x)$$

The mean  $\mu$  is also called the **expected value** of  $x$ ,  $E(x)$ .

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

We call  $\sigma$  the **standard deviation**.

We then studied a particular discrete random variable—the binomial random variable. This counts the number of successes in a **binomial experiment** consisting of  $n$  repeated independent trials or repetitions of some basic experiment, which has two possible outcomes called success ( $S$ ) and failure ( $F$ ). If the probability of success on a given trial is  $p$  and the probability of failure is  $q$ , the probability of exactly  $x$  successes in  $n$  trials is given by the binomial distribution

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

Finally, for the binomial random variable  $x$ ,

$$\begin{aligned}\mu &= n \cdot p \\ \sigma^2 &= n \cdot p \cdot q\end{aligned}$$

## REVIEW EXERCISES

- 6.56 A fair die is to be tossed. Let  $x$  = the number of tosses up to and including the first time a 6 appears.
- What values can  $x$  assume?
  - Is  $x$  a discrete or continuous random variable?
  - Is this a binomial experiment? Why?
- 6.57 A medical doctor is interested in studying the effect of a low-cholesterol diet on people with high blood cholesterol levels (measured in milligrams per 100 milliliters). To study this, she obtains 50 volunteers with high blood cholesterol levels to try the diet.
- Describe a random variable of possible interest to the medical doctor. Is it discrete or continuous?
  - Describe the population and sample of interest.

In Exercises 6.58–6.61, determine whether a probability distribution is given. Give a reason for your answer.

6.58

| $x$ | $P(x)$        |
|-----|---------------|
| 0   | $\frac{1}{4}$ |
| 4   | $\frac{3}{4}$ |

6.59

| $x$ | $P(x)$        |
|-----|---------------|
| 2   | $\frac{2}{3}$ |
| -4  | $\frac{1}{6}$ |
| 6   | $\frac{1}{6}$ |

6.60

| $x$ | $P(x)$        |
|-----|---------------|
| 0   | $\frac{1}{2}$ |
| 1   | $\frac{1}{2}$ |
| 2   | $\frac{1}{2}$ |

6.61

| $x$ | $P(x)$         |
|-----|----------------|
| 0   | $\frac{1}{5}$  |
| 1   | $\frac{2}{5}$  |
| 2   | $\frac{3}{5}$  |
| 3   | $\frac{1}{5}$  |
| 4   | $-\frac{2}{5}$ |

- 6.62 Determine (a) the mean, (b) the variance, and (c) the standard deviation of the following probability distribution:

| $x$ | $P(x)$        |
|-----|---------------|
| 2   | $\frac{4}{6}$ |
| 3   | $\frac{1}{6}$ |
| 13  | $\frac{1}{6}$ |

- 6.63 Determine (a)  $P(0)$ , (b) the mean, (c) the variance, and (d) the standard deviation of the following probability distribution:

| $x$ | $P(x)$        |
|-----|---------------|
| 0   | ?             |
| 2   | $\frac{1}{2}$ |
| 5   | $\frac{2}{5}$ |

- 6.64 Determine (a)  $P(3)$ , (b) the mean, (c) the variance, and (d) the standard deviation of the following probability distribution:

| $x$ | $P(x)$         |
|-----|----------------|
| 1   | $\frac{2}{12}$ |
| 2   | $\frac{5}{12}$ |
| 3   | ?              |
| 6   | $\frac{2}{12}$ |

- 6.65 Find the value of  $c$  so that a probability distribution is determined:

$$P(x) = \begin{cases} c, & \text{for } x = 1 \\ 2c, & \text{for } x = 2 \\ 3c, & \text{for } x = 3 \\ 4c, & \text{for } x = 4 \end{cases}$$

- 6.66 Find the value of  $c$  so that a probability distribution is determined:

$$P(x) = \begin{cases} c, & \text{for } x = 0 \\ 2c, & \text{for } x = 1 \end{cases}$$

- 6.67 Find the value of  $c$  so that a probability distribution is determined:

$$P(x) = \frac{cx}{6}, \quad \text{for } x = 1, 2$$

- 6.68 Three cards are numbered 1, 2, 3, and the experiment consists of turning the cards over one by one. A match occurs if the number of the card matches the number of the turn the card appears. So if the cards turn up in the order 2–1–3, there is one match, namely, the 3. Let  $x$  be the number of matches.

- (a) Find the probability distribution. (*Hint:* List the six possible outcomes.)  
 (b) If 60 people were to independently perform the experiment, what would you expect the approximate mean to be?

- 6.69 In the past year, 10% of RGH Airlines flights violated one or more safety codes. A government inspector decided to monitor one flight per day for the next 3 days. Let  $x$  be the number of flights that violate one or more safety codes. Find each of the following:

- (a) The probability distribution  
 (b) The mean of  $x$   
 (c) The variance of  $x$   
 (d) The standard deviation of  $x$

- 6.70 The probability distribution for the number of incoming phone calls  $x$  over a 5-minute period at a switchboard for a small business is

| $x$ | $P(x)$ |
|-----|--------|
| 0   | .3679  |
| 1   | .3679  |
| 2   | .1839  |
| 3   | .0613  |
| 4   | .0153  |
| 5   | .0031  |
| 6   | .0005  |
| 7   | .0001  |

In parts (a)–(c), find each of the following:

- (a) The mean of  $x$
- (b) The variance of  $x$
- (c) The standard deviation of  $x$
- (d) Construct a probability histogram.
- (e) Over a 5-minute period, what is the probability of one or more incoming phone calls?

6.71 A *binomial* probability distribution for the number of successes  $x$  is

| $x$ | $P(x)$ |
|-----|--------|
| 0   | .2401  |
| 1   | .4116  |
| 2   | .2646  |
| 3   | .0756  |
| 4   | .0081  |

In parts (a)–(c), find each of the following:

- (a) The mean of  $x$
- (b) The variance of  $x$
- (c) The standard deviation of  $x$
- (d) Construct a probability histogram.

[Hint: Find  $q$  from the fact that  $q^4 = P(0)$ . Now use the binomial shortcut formulas for  $\mu$  and  $\sigma^2$ .]

6.72 A *binomial* probability distribution for the number of successes  $x$  is

| $x$ | $P(x)$ |
|-----|--------|
| 0   | .36    |
| 1   | .48    |
| 2   | .16    |

find each of the following:

- (a) The mean of  $x$
- (b) The variance of  $x$
- (c) The standard deviation of  $x$
- (d) Construct a probability histogram.

[Hint: Use the binomial shortcut formulas and the fact that  $q^2 = P(0)$ .]

- 6.73 Match the three binomial probability histograms shown here with the correct  $p$ :  $p = .3$ ,  $p = .5$ ,  $p = .8$ .
- 6.74 A roulette wheel has 37 equally spaced slots that are numbered  $0, 1, 2, \dots, 36$ . Before the wheel is spun, a player picks one of the numbers. If the wheel is spun and stops on the number, the player wins \$35. If the wheel stops on one of the other numbers, the player pays \$1. What are the player's expected winnings (per game)?
- 6.75 In the game of craps, a player rolls two dice. If the first roll results in a sum of 7 or 11, the player wins. If the first roll results in a 2, 3, or 12, the player loses. If the sum on the first roll is 4, 5, 6, 8, 9, or 10, the player keeps rolling until he throws a 7 or the original value. If the outcome is a 7, the player loses. If it is the original value, the player wins. The probability that a player will win the game is .493. If the player nets \$4 for a win, what should he pay to play the game so that the game favors neither the player nor the casino?
- 6.76 A friend proposes the following game. He is to select one card from an ordinary deck of cards. You are to pay him \$10 if he selects an ace and \$5 if he selects a 2 or 3. What should he pay you so that the game does not favor either person?
- 6.77 The probability that a new flu vaccine prevents the flu is .80. Twelve people are vaccinated with the new vaccine at a local clinic. What is the probability that
- (a) Eight or more people will not get the flu?
  - (b) Two of the people will get the flu?
- 6.78 Assume that the probability of a boy being born is .50. If a couple plan on having six children, find the probability that
- (a) Exactly half are boys.
  - (b) All are boys.
  - (c) All are boys or all are girls.
  - (d) One or more are boys.
- (Assume that the gender of any child is independent of the genders of the other children.)
- 6.79 Thirty percent of all voters in a large city are Independents. If 15 voters are to be selected by chance, find the probability that
- (a) Exactly four are Independents.
  - (b) No more than three are Independents.
  - (c) Ten or more are Independents.
- 6.80 Seventy percent of the complaints received by a consumer protection agency are investigated by the agency. For the next seven complaints received by the agency, find the probability that
- (a) All will be investigated.
  - (b) One will be investigated.
  - (c) No more than two will be investigated.
- 6.81 For an electronic system to function, each of 12 components must work. Assume that the components function independently of each other, and each has probability .99 of not breaking down for 1 year, when new. Suppose you buy a new system. Find the probability that the system does not break down in the first year.

- 6.82 Eighty-five percent of the dishwashers manufactured by a large company do not need repairs for the first 2 years. If five of the dishwashers are to be selected by chance, find the probability that
- (a) None of the five will need repairs within 2 years.
  - (b) Three or more will need repairs within 2 years.

In Exercises 6.83–6.86, find (a) the mean, (b) the variance, and (c) the standard deviation of  $x$ .

- 6.83 Use Exercise 6.77 with  $x$  = the number of the 12 people who do not get the flu.  
 6.84 Use Exercise 6.80 with  $x$  = the number of complaints that will be investigated.  
 6.85 Use Exercise 6.79 with  $x$  = the number of the 15 voters who are Independent.  
 6.86 Use Exercise 6.82 with  $x$  = the number of dishwashers that will not need repairs in the next 2 years.

### ANSWERS TO PRACTICE QUIZ

1. .229
2. .047  
.254
3. .009
4. 2
5.
 

LET K1 = SUM(C1\*C2)  
 LET K2 = SUM(C1\*\*2\*C2) - K1\*\*2  
 LET K3 = SQRT(K2)  
 PRINT K1 - K3

(Note:  $K1 = \mu$ ,  $K2 = \sigma^2$ ,  $K3 = \sigma$ .)

### Notes

DeGroot, M. H., S. E. Fienberg, and J. B. Kadane, eds., *Statistics and the Law*. New York: John Wiley, 1986.

Kashy, T., *Finite Mathematics and Calculus with Applications*. Santa Monica, Calif.: Goodyear, 1979.

Larsen, R. J., and M. L. Marx, *Introduction to Mathematical Statistics and Its Applications*. Englewood Cliffs, N.J.: Prentice-Hall, 1986.

*The 1994 Information Please Almanac*. Boston: Houghton Mifflin, 1994.

Paulos, J. A. *Innumeracy—Mathematical Illiteracy and Its Consequences*. New York: Vintage Books, 1990.

**Figure 6.1**

**Figure 6.2**  
**A Probability Histogram**

**Figure 6.3**  
**Binomial Probability Distributions**