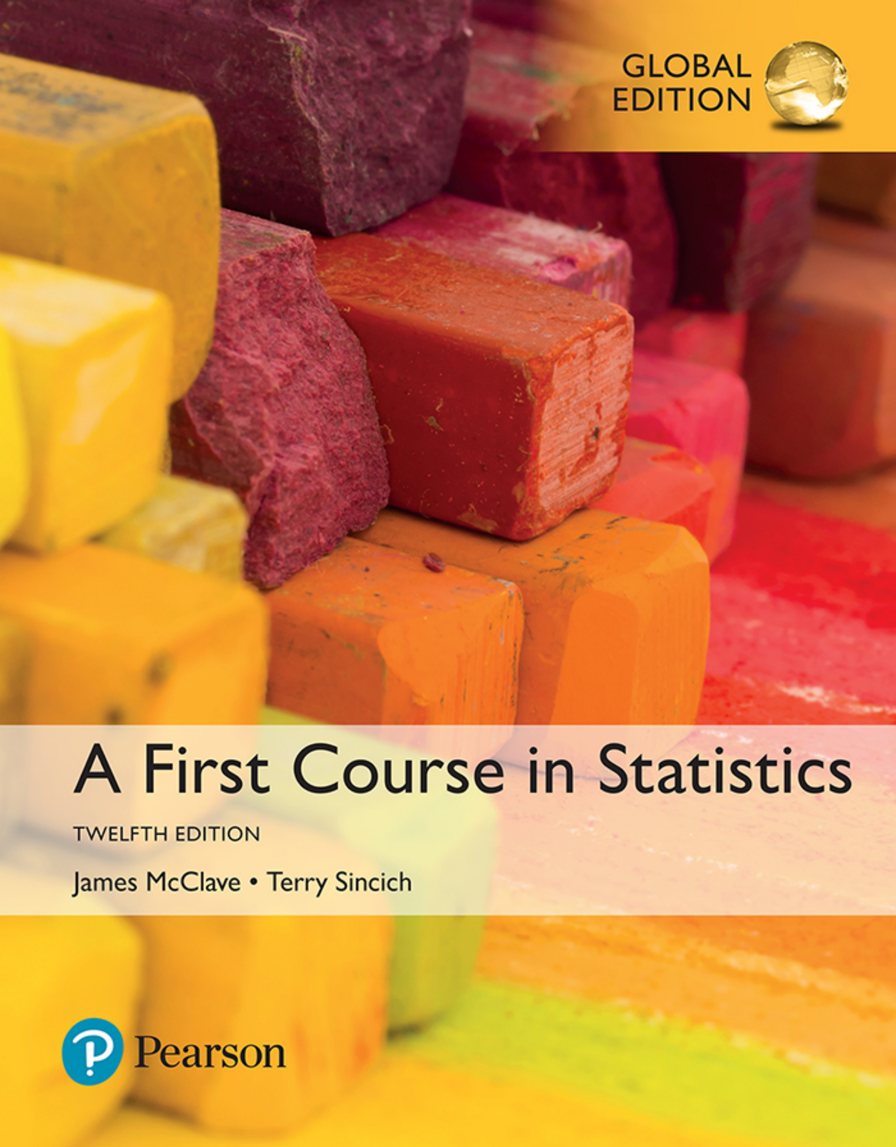

GLOBAL EDITION

A First Course in Statistics

TWELFTH EDITION

James McClave • Terry Sincich

Pearson

APPLET CORRELATION

Applet	Concept Illustrated	Description	Applet Activity
Sample from a population	Assesses how well a sample represents the population and the role that sample size plays in the process.	Produces random sample from population from specified sample size and population distribution shape. Reports mean, median, and standard deviation; applet creates plot of sample.	**4.4**, 216; **4.6**, 231
Sampling distributions	Compares means and standard deviations of distributions; assesses effect of sample size; illustrates unbiasedness.	Simulates repeatedly choosing samples of a fixed size *n* from a population with specified sample size, number of samples, and shape of population distribution. Applet reports means, medians, and standard deviations; creates plots for both.	**4.7,** 260; **4.8**, 260
Random numbers	Uses a random number generator to determine the experimental units to be included in a sample.	Generates random numbers from a range of integers specified by the user.	**1.1**, 43; **1.2**, 44; **3.6**, 183; **4.1**, 202
Long-run probability demonstrations illustrate the concept that theoretical probabilities are long-run experimental probabilities.			
Simulating probability of rolling a 6	Investigates relationship between theoretical and experimental probabilities of rolling 6 as number of die rolls increases.	Reports and creates frequency histogram for each outcome of each simulated roll of a fair die. Students specify number of rolls; applet calculates and plots proportion of 6s.	**3.1**, 151; **3.3**, 162; **3.4**, 163; **3.5**, 177
Simulating probability of rolling a 3 or 4	Investigates relationship between theoretical and experimental probabilities of rolling 3 or 4 as number of die rolls increases.	Reports outcome of each simulated roll of a fair die; creates frequency histogram for outcomes. Students specify number of rolls; applet calculates and plots proportion of 3s and 4s.	**3.3**, 162; **3.4**, 163
Simulating the probability of heads: fair coin	Investigates relationship between theoretical and experimental probabilities of getting heads as number of fair coin flips increases.	Reports outcome of each fair coin flip and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots proportion of heads.	**3.2**, 151; **4.2**, 203
Simulating probability of heads: unfair coin ($P(H) = .2$)	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is less likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	**4.3**, 216
Simulating probability of heads: unfair coin ($P(H) = .8$)	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is more likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	**4.3**, 216
Simulating the stock market	Theoretical probabilities are long run experimental probabilities.	Simulates stock market fluctuation. Students specify number of days; applet reports whether stock market goes up or down daily and creates a bar graph for outcomes. Calculates and plots proportion of simulated days stock market goes up.	**4.5**, 216
Mean versus median	Investigates how skewedness and outliers affect measures of central tendency.	Students visualize relationship between mean and median by adding and deleting data points; applet automatically updates mean and median.	**2.1**, 85; **2.2**, 85; **2.3**, 85

Example 4.2

Values of a Discrete Random Variable—EPA Application

Problem Suppose the Environmental Protection Agency (EPA) takes readings once a month on the amount of pesticide in the discharge water of a chemical company. If the amount of pesticide exceeds the maximum level set by the EPA, the company is forced to take corrective action and may be subject to penalty. Consider the random variable x to be the number of months before the company's discharge exceeds the EPA's maximum level. What values can x assume?

Solution The company's discharge of pesticide may exceed the maximum allowable level on the first month of testing, the second month of testing, etc. It is possible that the company's discharge will *never* exceed the maximum level. Thus, the set of possible values for the number of months until the level is first exceeded is the set of all positive integers 1, 2, 3, 4, ...

Look Back If we can list the values of a random variable x, even though the list is never ending, we call the list **countable** and the corresponding random variable **discrete**. Thus, the number of months until the company's discharge first exceeds the limit is a **discrete random variable**.

■ **Now Work Exercise 4.7**

Example 4.3

Values of a Continuous Random Variable—Another EPA Application

Problem Refer to Example 4.2. A second random variable of interest is the amount x of pesticide (in milligrams per liter) found in the monthly sample of discharge waters from the same chemical company. What values can this random variable assume?

Solution Some possible values of x are 1.7, 28.42, and 100.987 milligrams per liter. Unlike the *number* of months before the company's discharge exceeds the EPA's maximum level, the set of all possible values for the *amount* of discharge *cannot* be listed (i.e., is not countable). The possible values for the amount x of pesticide would correspond to the points on the interval between 0 and the largest possible value the amount of the discharge could attain, the maximum number of milligrams that could occupy 1 liter of volume. (Practically, the interval would be much smaller, say, between 0 and 500 milligrams per liter.)

Look Ahead When the values of a random variable are not countable but instead correspond to the points on some interval, we call the variable a *continuous random variable*. Thus, the *amount* of pesticide in the chemical plant's discharge waters is a *continuous random variable*.

■ **Now Work Exercise 4.9**

> Random variables that can assume a *countable* number of values are called **discrete**.

> Random variables that can assume values corresponding to any of the points contained in an interval are called **continuous**.

The following are examples of discrete random variables:

1. The number of seizures an epileptic patient has in a given week: $x = 0, 1, 2, \ldots$
2. The number of voters in a sample of 500 who favor impeachment of the president: $x = 0, 1, 2, \ldots, 500$
3. The shoe size of a tennis player: $x = \ldots\ 5, 5\frac{1}{2}, 6, 6\frac{1}{2}, 7, 7\frac{1}{2} \ldots$
4. The change received for paying a bill: $x =$ 1¢, 2¢, 3¢, . . . , \$1, \$1.01, \$1.02, . . .
5. The number of customers waiting to be served in a restaurant at a particular time: $x = 0, 1, 2, \ldots$

Note that several of the examples of discrete random variables begin with the words *The number of.* ... This wording is very common, since the discrete random variables most frequently observed are counts. The following are examples of continuous random variables:

1. The length of time (in seconds) between arrivals at a hospital clinic: $0 \leq x \leq \infty$ (infinity)
2. The length of time (in minutes) it takes a student to complete a one-hour exam: $0 \leq x \leq 60$
3. The amount (in ounces) of carbonated beverage loaded into a 12-ounce can in a can-filling operation: $0 \leq x \leq 12$
4. The depth (in feet) at which a successful oil-drilling venture first strikes oil: $0 \leq x \leq c$, where c is the maximum depth obtainable
5. The weight (in pounds) of a food item bought in a supermarket: $0 \leq x \leq 500$ [*Note:* Theoretically, there is no upper limit on x, but it is unlikely that it would exceed 500 pounds.]

Discrete random variables and their probability distributions are discussed in Sections 4.2 and 4.3. Continuous random variables and their probability distributions are the topic of Sections 4.4 and 4.5.

Exercises 4.1–4.16

Understanding the Principles

4.1 Define a discrete random variable.

4.2 Define a continuous random variable.

Applying the Concepts—Basic

4.3 **Type of Random Variable.** Classify the following random variables according to whether they are discrete or continuous:

a. The number of students in a class
b. The height of students in a college
c. The proportion of girls in a class of 100 students
d. The monthly income of employees in a company
e. The water flowing through the Hirakud Dam
f. The amount of rainfall in a particular city.

4.4 **Type of Random Variable.** Identify the following random variables as discrete or continuous:

a. The life span of an electric bulb
b. The heart rate (number of beats per minute) of an adult male
c. The number of defective push pins produced by a machine in a company
d. The temperature in a particular city
e. The population of fishes in a pond.
f. The number of leaks in a 100 kilometer oil pipeline

4.5 **Type of Random Variable.** Identify the following variables as discrete or continuous:

a. The time span of failure of an electronic component
b. The number of accidents committed per month in your community
c. The number of machine breakdowns during a given day
d. The number of matches won by a player
e. The number of faulty blades in a pack of 100
f. The amount of water in a bottle

4.6 **NHTSA crash tests.** The National Highway Traffic Safety Administration (NHTSA) has developed a driver-side "star" scoring system for crash-testing new cars. Each crash-tested car is given a rating ranging from one star (*) to five stars (*****): the more stars in the rating, the better is the level of crash protection in a head-on collision. Suppose that a car is selected and its driver-side star rating is determined. Let x equal the number of stars in the rating. Is x a discrete or continuous random variable?

4.7 NW **Customers in line at a Subway shop.** The number of customers, x, waiting in line to order sandwiches at a Subway shop at noon is of interest to the store manager. What values can x assume? Is x a discrete or continuous random variable?

4.8 **Sound waves from a basketball.** Refer to the *American Journal of Physics* (June 2010) experiment on sound waves produced from striking a basketball, Exercise 2.43 (p. 76). Recall that the frequencies of sound wave echoes resulting from striking a hanging basketball with a metal rod were recorded. Classify the random variable, frequency (measured in hertz) of an echo, as discrete or continuous.

4.9 NW **Mongolian desert ants.** Refer to the *Journal of Biogeography* (Dec. 2003) study of ants in Mongolia, presented in Exercise 2.68 (p. 87). Two of the several variables recorded at each of 11 study sites were annual rainfall (in millimeters) and number of ant species. Identify these variables as discrete or continuous.

4.10 Motivation of drug dealers. Refer to the *Applied Psychology in Criminal Justice* (Sept. 2009) study of the personality characteristics of drug dealers, Exercise 2.102 (p. 101). For each of 100 convicted drug dealers, the researchers measured several variables, including the number of prior felony arrests x. Is x a discrete or continuous random variable? Explain.

Applying the Concepts—Intermediate

4.11 Nutrition. Give an example of a discrete random variable of interest to a nutritionist.

4.12 Physics. Give an example of a discrete random variable of interest to a physicist.

4.13 Medical. Give an example of a discrete random variable of interest to a doctor in a hospital.

4.14 Artist. Give an example of a discrete random variable of interest to an artist.

4.15 Irrelevant speech effects. Refer to the *Acoustical Science & Technology* (Vol. 35, 2014) study of the degree to which the memorization process is impaired by irrelevant background speech (called irrelevant speech effects), Exercise 2.34 (p. 73). Recall that subjects performed a memorization task under two conditions: (1) with irrelevant background speech and (2) in silence. Let x represent the difference in the error rates for the two conditions—called the relative difference in error rate (RDER). Explain why x is a continuous random variable.

4.16 Shaft graves in ancient Greece. Refer to the *American Journal of Archaeology* (Jan. 2014) study of shaft graves in ancient Greece, Exercise 2.37 (p. 74). Let x represent the number of decorated sword shafts buried at a discovered grave site. Explain why x is a discrete random variable.

4.2 Probability Distributions for Discrete Random Variables

A complete description of a discrete random variable requires that we *specify all the values the random variable can assume* and *the probability associated with each value*. To illustrate, consider Example 4.4.

Example 4.4

Finding a Probability Distribution—Coin-Tossing Experiment

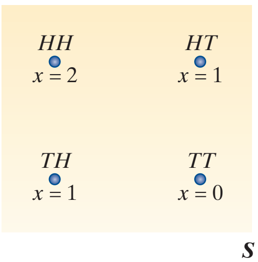

Figure 4.2
Venn diagram for the two-coin-toss experiment

Problem Recall the experiment of tossing two coins (p. 142), and let x be the number of heads observed. Find the probability associated with each value of the random variable x, assuming that the two coins are fair.

Solution The sample space and sample points for this experiment are reproduced in Figure 4.2. Note that the random variable x can assume values 0, 1, 2. Recall (from Chapter 3) that the probability associated with each of the four sample points is $^1/_4$. Then, identifying the probabilities of the sample points associated with each of these values of x, we have

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

$$P(x = 1) = P(TH) + P(HT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

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

Thus, we now know the values the random variable can assume (0, 1, 2) and how the probability is *distributed over* those values ($^1/_4$, $^1/_2$, $^1/_4$). This dual specification completely describes the random variable and is referred to as the *probability distribution*, denoted by the symbol $p(x)$.* The probability distribution for the coin-toss example is shown in tabular form in Table 4.1 and in graphic form in Figure 4.3. Since the probability distribution for a discrete random variable is concentrated at specific points (values of x), the graph in Figure 4.3a represents the probabilities as the heights of vertical lines over the corresponding values of x. Although the representation of the probability distribution as a histogram, as in Figure 4.3b, is less precise (since the probability is spread over a unit interval), the histogram representation will prove useful when we approximate probabilities of certain discrete random variables in Section 4.4.

*In standard mathematical notation, the probability that a random variable X takes on a value x is denoted $P(X = x) = p(x)$. Thus, $P(X = 0) = p(0)$, $P(X = 1) = p(1)$, etc. In this text, we adopt the simpler $p(x)$ notation.

Table 4.1 Probability Distribution for Coin-Toss Experiment: Tabular Form

x	$p(x)$
0	$1/4$
1	$1/2$
2	$1/4$

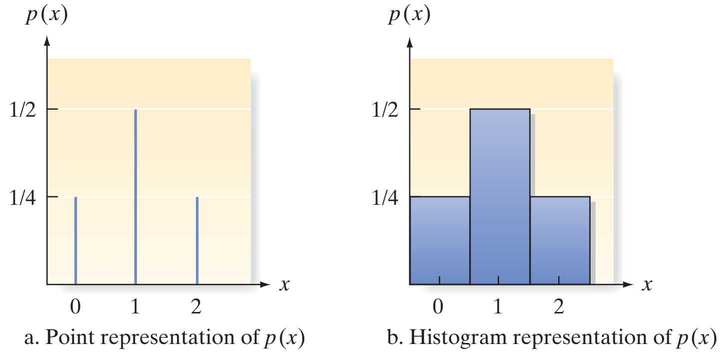

Figure 4.3
Probability distribution for coin-toss experiment: graphical form

Look Ahead We could also present the probability distribution for x as a formula, but this would unnecessarily complicate a very simple example. We give the formulas for the probability distributions of some common discrete random variables later in the chapter.

■ **Now Work Exercise 4.24**

The **probability distribution of a discrete random variable** is a graph, table, or formula that specifies the probability associated with each possible value that the random variable can assume.

Two requirements must be satisfied by all probability distributions for discrete random variables:

Requirements for the Probability Distribution of a Discrete Random Variable *x*

1. $p(x) \geq 0$ for all values of x.

2. $\sum p(x) = 1$

where the summation of $p(x)$ is over all possible values of x.*

Example 4.5

Probability Distribution from a Graph—Playing Craps

Problem Craps is a popular casino game in which a player throws two dice and bets on the outcome (the sum total of the dots showing on the upper faces of the two dice). Consider a \$5 wager. On the first toss (called the *come-out* roll), if the total is 7 or 11 the roller wins \$5. If the outcome is a 2, 3, or 12, the roller loses \$5 (i.e., the roller wins $-\$5$). For any other outcome (4, 5, 6, 8, 9, or 10), a *point* is established and no money is lost or won on that roll (i.e., the roller wins \$0). In a computer simulation of repeated tosses of two dice, the outcome x of the come-out roll wager ($-\$5$, \$0, or $+\$5$) was recorded. A relative frequency histogram summarizing the results is shown in Figure 4.4. Use the histogram to find the approximate probability distribution of x.

Solution The histogram shows that the relative frequencies of the outcomes $x = -\$5$, $x = \$0$, and $x = \$5$ are .1, .65, and .25, respectively. For example, in repeated tosses of two dice, 25% of the outcomes resulted in a sum of 7 or 11 (a \$5 win for the roller). Based on our long-run definition of probability given in Chapter 3, these relative frequencies estimate the probabilities of the three outcomes. Consequently, the approximate probability distribution of x, the outcome of the come-out wager in craps, is $p(-\$5) = .1$, $p(\$0) = .65$, and $p(\$5) = .25$. Note that these probabilities sum to 1.

*Unless otherwise indicated, summations will always be over all possible values of x.

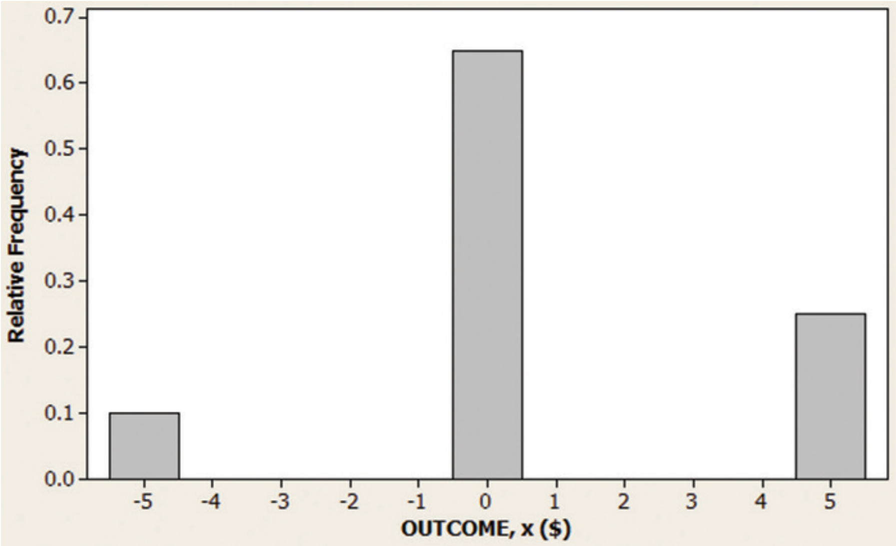

Figure 4.4
MINITAB Histogram for \$5 Wager on Come-Out Roll in Craps

Look Back When two dice are tossed, there is a total of 36 possible outcomes. (Can you list these outcomes, or sample points?) Of these, 4 result in a sum of 2, 3, or 12; 24 result in a sum of 4, 5, 6, 8, 9, or 10; and 8 result in a sum of 7 or 11. Using the rules of probability established in Chapter 3, you can show that the actual probability distribution for x is $p(-\$5) = 4/36 = .1111$, $p(\$0) = 24/36 = .6667$, and $p(\$5) = 8/36 = .2222$.

■ **Now Work Exercise 4.21**

Examples 4.4 and 4.5 illustrate how the probability distribution for a discrete random variable can be derived, but for many practical situations the task is much more difficult. Fortunately, numerous experiments and associated discrete random variables observed in nature possess identical characteristics. Thus, you might observe a random variable in a psychology experiment that would possess the same probability distribution as a random variable observed in an engineering experiment or a social sample survey. We classify random variables according to type of experiment, derive the probability distribution for each of the different types, and then use the appropriate probability distribution when a particular type of random variable is observed in a practical situation. The probability distributions for most commonly occurring discrete random variables have already been derived. This fact simplifies the problem of finding the probability distributions for random variables, as the next example illustrates.

Example 4.6

Probability Distribution Using a Formula—Texas Droughts

Problem A drought is a period of abnormal dry weather that causes serious problems in the farming industry of the region. University of Arizona researchers used historical annual data to study the severity of droughts in Texas (*Journal of Hydrologic Engineering*, Sept./Oct. 2003). The researchers showed that the distribution of x, the number of consecutive years that must be sampled until a dry (drought) year is observed, can be modeled using the formula

$$p(x) = (.3)(.7)^{x-1}, x = 1, 2, 3, \ldots$$

Find the probability that exactly 3 years must be sampled before a drought year occurs.

Solution We want to find the probability that $x = 3$. Using the formula, we have

$$p(3) = (.3)(.7)^{3-1} = (.3)(.7)^2 = (.3)(.49) = .147$$

Thus, there is about a 15% chance that exactly 3 years must be sampled before a drought year occurs in Texas.

Look Back The probability of interest can also be derived using the principles of probability developed in Chapter 3. The event of interest is $N_1N_2D_3$, where N_1 represents no drought occurs in the first sampled year, N_2 represents no drought occurs in the second sampled year, and D_3 represents a drought occurs in the third sampled year. The researchers discovered that the probability of a drought occurring in any sampled year is .3 (and, consequently, the probability of no drought occurring in any sampled year is .7). Using the multiplicative rule of probability for independent events, the probability of interest is $(.7)(.7)(.3) = .147$.

■ **Now Work Exercise 4.36**

Since probability distributions are analogous to the relative frequency distributions of Chapter 2, it should be no surprise that the mean and standard deviation are useful descriptive measures.

If a discrete random variable x were observed a very large number of times and the data generated were arranged in a relative frequency distribution, the relative frequency distribution would be indistinguishable from the probability distribution for the random variable. Thus, the probability distribution for a random variable is a theoretical model for the relative frequency distribution of a population. To the extent that the two distributions are equivalent (and we will assume that they are), the probability distribution for x possesses a mean μ and a variance σ^2 that are identical to the corresponding descriptive measures for the population. The procedure for finding μ and σ^2 of a random variable follows.

Examine the probability distribution for x (the number of heads observed in the toss of two fair coins) in Figure 4.5. Try to locate the mean of the distribution intuitively. We may reason that the mean μ of this distribution is equal to 1 as follows: In a large number of experiments—say, 100,000—$^1/_4$ (or 25,000) should result in $x = 0$ heads, $^1/_2$ (or 50,000) in $x = 1$ head, and $^1/_4$ (or 25,000) in $x = 2$ heads. Therefore, the average number of heads is

$$\mu = \frac{0(25{,}000) + 1(50{,}000) + 2(25{,}000)}{100{,}000} = 0(^1/_4) + 1(^1/_2) + 2(^1/_4)$$
$$= 0 + {}^1/_2 + {}^1/_2 = 1$$

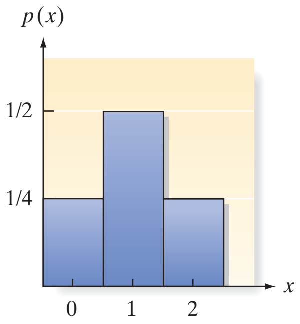

Figure 4.5
Probability distribution for a two-coin toss

Note that to get the population mean of the random variable x, we multiply each possible value of x by its probability $p(x)$, and then we sum this product over all possible values of x. The *mean of x* is also referred to as the *expected value of x*, denoted $E(x)$.

The **mean**, or **expected value**, of a discrete random variable x is

$$\mu = E(x) = \Sigma xp(x)$$

Expected is a mathematical term and should not be interpreted as it is typically used. Specifically, *a random variable might never be equal to its "expected value."* Rather, the expected value is the mean of the probability distribution, or a measure of its central tendency. You can think of μ as the mean value of x in a *very large* (actually, *infinite*) number of repetitions of the experiment in which the values of x occur in proportions equivalent to the probabilities of x.

Example 4.7

Finding an Expected Value—An Insurance Application

Problem Suppose you work for an insurance company and you sell a $10,000 one-year term insurance policy at an annual premium of $290. Actuarial tables show that the probability of death during the next year for a person of your customer's age, sex, health, etc., is .001. What is the expected gain (amount of money made by the company) for a policy of this type?

Solution The experiment is to observe whether the customer survives the upcoming year. The probabilities associated with the two sample points, Live and Die, are .999 and .001,