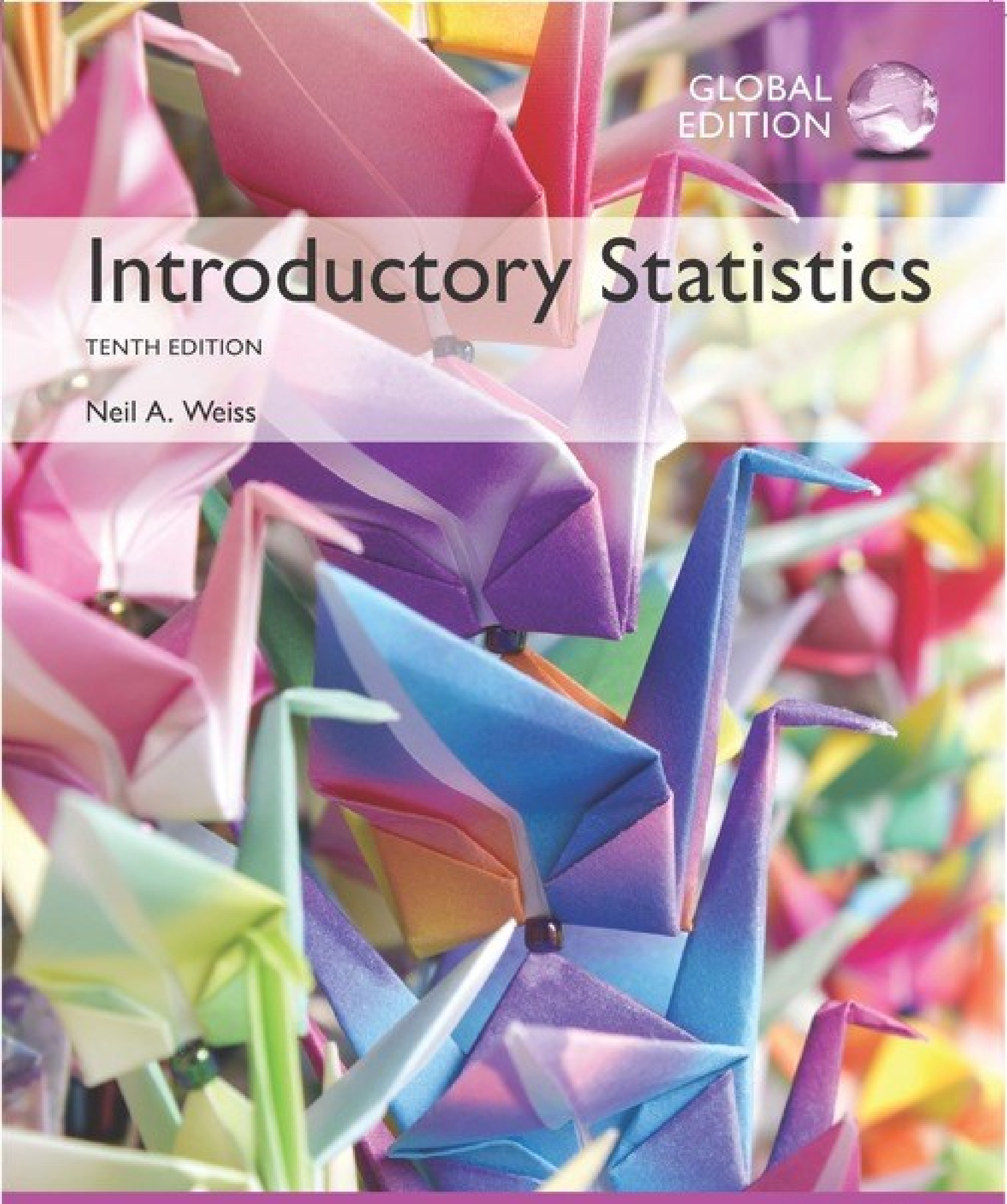
GLOBAL EDITION

Introductory Statistics

TENTH EDITION

Neil A. Weiss



ALWAYS LEARNING

PEARSON

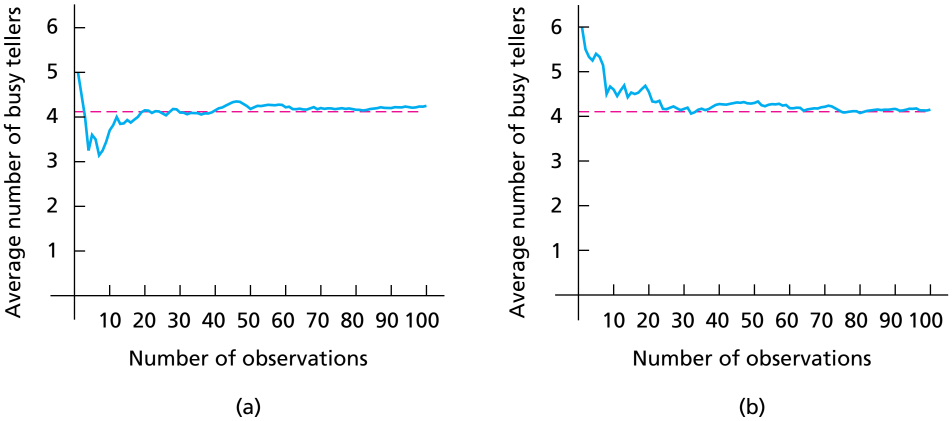
Introductory

STATISTICS

10TH EDITION

GLOBAL EDITION

FIGURE 5.3
Graphs showing the average number of busy tellers versus the number of observations for two simulations of 100 observations each



DEFINITION 5.5

Standard Deviation of a Discrete Random Variable

The **standard deviation of a discrete random variable X** is denoted $\boldsymbol{\sigma_X}$ or, when no confusion will arise, simply $\boldsymbol{\sigma}$. It is defined as

$$\sigma = \sqrt{\Sigma(x-\mu)^2 P(X=x)}.$$

The standard deviation of a discrete random variable can also be obtained from the computing formula

$$\sigma = \sqrt{\Sigma x^2 P(X=x) - \mu^2}.$$

What Does It Mean?

Roughly speaking, the standard deviation of a random variable X indicates how far, on average, an observed value of X is from its mean. In particular, the smaller the standard deviation of X, the more likely it is that an observed value of X will be close to its mean.

Note: The square of the standard deviation, $\boldsymbol{\sigma^2}$, is called the **variance** of X.

EXAMPLE 5.8 The Standard Deviation of a Discrete Random Variable

Busy Tellers Recall Example 5.7, where X denotes the number of tellers busy with customers at 1:00 P.M. Find the standard deviation of X.

Solution We apply the computing formula given in Definition 5.5. To use that formula, we need the mean of X, which we found in Example 5.7 to be 4.118, and columns for x^2 and $x^2P(X=x)$, which are presented in the last two columns of Table 5.12.

TABLE 5.12
Table for computing the standard deviation of the random variable X, the number of tellers busy with customers

x	$P(X=x)$	x^2	$x^2P(X=x)$
0	0.029	0	0.000
1	0.049	1	0.049
2	0.078	4	0.312
3	0.155	9	1.395
4	0.212	16	3.392
5	0.262	25	6.550
6	0.215	36	7.740
			19.438

From the final column of Table 5.12, $\Sigma x^2 P(X=x) = 19.438$. Thus

$$\sigma = \sqrt{\Sigma x^2 P(X=x) - \mu^2} = \sqrt{19.438 - (4.118)^2} = 1.6.$$

Interpretation Roughly speaking, on average, the number of busy tellers is 1.6 from the mean of 4.118 busy tellers.

You try it!

Exercise 5.35(b) on page 257

Exercises 5.2

Understanding the Concepts and Skills

5.25 What concept does the mean of a discrete random variable generalize?

5.26 Comparing Investments. Suppose that the random variables X and Y represent the amount of return on two different investments. Further suppose that the mean of X equals the mean of Y but that the standard deviation of X is greater than the standard deviation of Y.

a. On average, is there a difference between the returns of the two investments? Explain your answer.

b. Which investment is more conservative? Why?

*In Exercises **5.27–5.30**, we have provided the probability distributions of the random variables considered in Exercises 5.7–5.10 of Section 5.1. For each exercise, do the following tasks.*

a. *Find the mean of the random variable.*

b. *Obtain the standard deviation of the random variable by using one of the formulas given in Definition 5.5.*

5.27

x	1	2	3
$P(X = x)$	0.3	0.4	0.3

5.28

y	1	2	3	4
$P(Y = y)$	0.1	0.1	0.5	0.3

5.29

y	0	1	4	6
$P(Y = y)$	0.36	0.28	0.16	0.20

5.30

x	5	7	8
$P(X = x)$	0.20	0.32	0.48

Applying the Concepts and Skills

*In Exercises **5.31–5.35**, we have provided the probability distributions of the random variables considered in Exercises 5.11–5.15 of Section 5.1. For each exercise, do the following tasks.*

a. *Find and interpret the mean of the random variable.*

b. *Obtain the standard deviation of the random variable by using one of the formulas given in Definition 5.5.*

c. *Construct a probability histogram for the random variable; locate the mean; and show one-, two-, and three-standard-deviation intervals.*

5.31 Space Shuttles. The random variable X is the crew size of a randomly selected shuttle mission between April 12, 1981 and July 8, 2011. Its probability distribution is as follows.

x	2	4	5	6	7	8
$P(X = x)$	0.030	0.022	0.267	0.207	0.467	0.007

5.32 Persons per Housing Unit. The random variable Y is the number of persons living in a randomly selected occupied housing unit. Its probability distribution is as follows.

y	1	2	3	4	5	6	7
$P(Y = y)$	0.265	0.327	0.161	0.147	0.065	0.022	0.013

5.33 Major Hurricanes. The random variable Y is the number of major hurricanes for a randomly selected year between 1851 and 2012. Its probability distribution is as follows.

y	$P(Y = y)$	y	$P(Y = y)$
0	0.185	5	0.056
1	0.296	6	0.037
2	0.266	7	0.012
3	0.093	8	0.006
4	0.049		

5.34 Accident. Let X be the random variable representing number of motorcycle accidents that has happened in a region per second. Its probability distribution is as follows.

x	0	1	2	3
$P(X = x)$	0.125	0.375	0.375	0.125

5.35 Dice. The random variable Y is the sum of the dice when two balanced dice are rolled. Its probability distribution is as follows.

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y = y)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

5.36 World Series. The World Series in baseball is won by the first team to win four games (ignoring the 1903 and 1919–1921 World Series, when it was a best of nine). From the document *World Series History* on the Baseball Almanac website, as of November 2013, the lengths of the World Series are as given in the following table.

Number of games	Frequency	Relative frequency
4	21	0.200
5	24	0.229
6	24	0.229
7	36	0.343

Let X denote the number of games that it takes to complete a World Series, and let Y denote the number of games that it took to complete a randomly selected World Series from among those considered in the table.

a. Determine the mean and standard deviation of the random variable Y. Interpret your results.

b. Provide an estimate for the mean and standard deviation of the random variable X. Explain your reasoning.

5.37 10 Air Meter Pistol Event. In a 10-Metre-Air-Pistol event, a 4.5 mm caliber air gun is shot from a distance of 10 meters into a circular target with a 6 feet radius whose center we call the origin. The program consists of 60 shots within 105 minutes for men. The outcome of this random experiment is a shot on the target. The shooter scores 10 points if he hits the bull's eye, which is a disk with radius of 1-foot centered at the origin; he scores 5 points if he hits the ring with inner radius of 1 foot and outer radius of 3 feet centered at the origin; and he scores 0 points if he shoots anywhere outside. Assume that the shooter will actually hit the target. For one shot, let S be the score. A probability distribution for the random variable S is as follows.

s	0	5	10
$P(S = s)$	0.7611	0.2112	0.0277

a. On average, how many points will the shooter score per shot?
b. Obtain and interpret the standard deviation of the score per shot.

5.38 All-Numeric Passwords. The technology consultancy Data-Genetics published the online document *PIN analysis*. In addition to analyzing PIN numbers, passwords trends were examined. Seven million all-numeric passwords were collected and yielded the following estimate of the probability distribution of the number of digits used in an all-numeric password.

Digits	Probability	Digits	Probability
4	0.492	9	0.029
5	0.092	10	0.015
6	0.178	11	0.004
7	0.073	12	0.003
8	0.113	13	0.001

On average, approximately how many digits would you expect an all-numeric password to have? Explain your answer.

Expected Value. *As noted in Definition 5.4 on page 254, the mean of a random variable is also called its expected value. This terminology is especially useful in gambling, decision theory, and the insurance industry, as illustrated in Exercises 5.39–5.42.*

5.39 Roulette. An American roulette wheel contains 38 numbers: 18 are red, 18 are black, and 2 are green. When the roulette wheel is spun, the ball is equally likely to land on any of the 38 numbers. Suppose that you bet \$1 on red. If the ball lands on a red number, you win \$1; otherwise you lose your \$1. Let X be the amount you win on your \$1 bet. Then X is a random variable whose probability distribution is as follows.

x	1	−1
$P(X = x)$	0.474	0.526

a. Verify that the probability distribution is correct.
b. Find the expected value of the random variable X.
c. On average, how much will you lose per play?
d. Approximately how much would you expect to lose if you bet \$1 on red 100 times? 1000 times?
e. Is roulette a profitable game to play? Explain.

5.40 Evaluating Investments. An investor plans to put \$50,000 in one of four investments. The return on each investment depends on whether next year's economy is strong or weak. The following table summarizes the possible payoffs, in dollars, for the four investments.

Investment	Next year's economy: Strong	Next year's economy: Weak
Certificate of deposit	6,000	6,000
Office complex	15,000	5,000
Land speculation	33,000	−17,000
Technical school	5,500	10,000

Let V, W, X, and Y denote the payoffs for the certificate of deposit, office complex, land speculation, and technical school, respectively. Then V, W, X, and Y are random variables. Assume that next year's economy has a 40% chance of being strong and a 60% chance of being weak.
a. Find the probability distribution of each random variable V, W, X, and Y.
b. Determine the expected value of each random variable.
c. Which investment has the best expected payoff? the worst?
d. Which investment would you select? Explain.

5.41 Life of AA Battery. A company wants to design an AA battery that lasts longer than the existing batteries. The data compiled by the company gives the life of the batteries, X, in minutes. We are assuming X as a random variable with the following probability distribution.

x	360	380	400	420	440
$P(X = x)$	0.148	0.139	0.572	0.123	0.018

a. Determine the expected minutes of the AA batteries.
b. How many minutes in total should the company aim for in order to increase battery life if it wants the average life to increase by 20 minutes?

5.42 Expected Utility. One method for deciding among various investments involves the concept of **expected utility.** Economists describe the importance of various levels of wealth by using **utility functions.** For instance, in most cases, a single dollar is more important (has greater utility) for someone with little wealth than for someone with great wealth. Consider two investments, say, Investment A and Investment B. Measured in thousands of dollars, suppose that Investment A yields 0, 1, and 4 with probabilities 0.1, 0.5, and 0.4, respectively, and that Investment B yields 0, 1, and 16 with probabilities 0.5, 0.3, and 0.2, respectively. Let Y denote the yield of an investment. For the two investments, determine and compare
a. the mean of Y, the expected yield.
b. the mean of $\sqrt{Y}$, the expected utility, using the utility function $u(y) = \sqrt{y}$. Interpret the utility function u.
c. the mean of $Y^{3/2}$, the expected utility, using the utility function $v(y) = y^{3/2}$. Interpret the utility function v.

5.43 Bulbs. A distributor orders 100 boxes of bulbs. A box of 10 flashbulbs contains 3 defective bulbs. A random sample of 2 bulbs is selected and tested. The probability distribution is shown in the following table, where X denotes the number of defective bulbs.

x	0	1	2
$P(X = x)$	0.181	0.459	0.36

a. Determine μ_W and σ_W. Round your answer for the standard deviation to three decimal places.
b. On average, how many defective bulbs are there in a box?
c. About how many defective bulbs are expected in 100 boxes?

Extending the Concepts and Skills

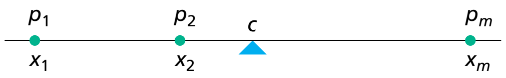
5.44 Simulation. Let X be the value of a randomly selected decimal digit, that is, a whole number between 0 and 9, inclusive.
a. Use simulation to estimate the mean of X. Explain your reasoning.
b. Obtain the exact mean of X by applying Definition 5.4 on page 254. Compare your result with that in part (a).

5.45 Queuing Simulation. Benny's Barber Shop in Cleveland has five chairs for waiting customers. The number of customers waiting is a random variable Y with the following probability distribution.

y	0	1	2	3	4	5
***P*(*Y* = *y*)**	0.424	0.161	0.134	0.111	0.093	0.077

a. Compute and interpret the mean of the random variable Y.
b. In a large number of independent observations, how many customers will be waiting, on average?
c. Use the technology of your choice to simulate 500 observations of the number of customers waiting.
d. Obtain the mean of the observations in part (c) and compare it to μ_Y.
e. What does part (d) illustrate?

5.46 Mean as Center of Gravity. Let X be a discrete random variable with a finite number of possible values, say, x_1, $x_2, \ldots, x_m$. For convenience, set $p_k = P(X = x_k)$, for $k = 1$, $2, \ldots, m$. Think of a horizontal axis as a seesaw and each p_k as a mass placed at point x_k on the seesaw. The *center of gravity* of these masses is defined to be the point c on the horizontal axis at which a fulcrum could be placed to balance the seesaw.



Relative to the center of gravity, the torque acting on the seesaw by the mass p_k is proportional to the product of that mass with the signed distance of the point x_k from c, that is, to $(x_k - c) \cdot p_k$. Show that the center of gravity equals the mean of the random variable X. (*Hint:* To balance, the total torque acting on the seesaw must be 0.)

Properties of the Mean and Standard Deviation. In Exercises 5.47 and 5.48, you will develop some important properties of the mean and standard deviation of a random variable. Two of them relate the mean and standard deviation of the sum of two random variables to the individual means and standard deviations, respectively; two others relate the mean and standard deviation of a constant times a random variable to the constant and the mean and standard deviation of the random variable, respectively.

In developing these properties, you will need to use the concept of independent random variables. Two discrete random variables, X and Y, are said to be *independent random variables* if

$$P(\{X = x\} \,\&\, \{Y = y\}) = P(X = x) \cdot P(Y = y)$$

for all x and y—that is, if the joint probability distribution of X and Y equals the product of their marginal probability distributions. This condition is equivalent to requiring that events $\{X = x\}$ and $\{Y = y\}$ are independent for all x and y. A similar definition holds for independence of more than two discrete random variables.

5.47 Bulbs. Refer to Exercise 5.43. Assume that the number of defective bulbs on different boxes are independent of one another. Let X and Y denote the number of defective bulbs on each of two boxes.

		y			
		0	**1**	**2**	***P*(*X* = *x*)**
x	0				
	1				
	2				
	***P*(*Y* = *y*)**				

a. Complete the preceding joint probability distribution table. *Hint:* To obtain the joint probability in the first row, third column, use the definition of independence for discrete random variables and the table in Exercise 5.43:

$$\begin{aligned} P(\{X = 0\} \,\&\, \{Y = 2\}) &= P(X = 0) \cdot P(Y = 2) \\ &= 0.181 \cdot 0.36 = 0.065. \end{aligned}$$

b. Use the joint probability distribution you obtained in part(a) to determine the probability distribution of the random variable $X + Y$, the total number of defective bulbs in two boxes; that is, complete the following table.

u	0	1	2	3	4
P*(*X* + *Y*) = *u					

c. Use part (b) to find μ_{X+Y} and σ^2_{X+Y}.
d. Use part (c) to verify that the following equations hold for this example:

$$\mu_{X+Y} = \mu_X + \mu_Y \qquad \sigma_{X+Y} = \sigma_X + \sigma_Y.$$

(*Note:* The mean and variance of X and Y are the same as that of X in Exercise 5.43.)
e. The equations in part (d) hold in general: If X and Y are any two random variables,

$$\mu_{X+Y} = \mu_X + \mu_Y.$$

In addition, if X and Y are independent,

$$\sigma^2_{X+Y} = \sigma^2_Y + \sigma^2_X.$$

Interpret these two equations in words.

5.48 Bulbs. The distributor in Exercise 5.43 estimates that each defective bulb costs him \$5 for replacement. If X is the number of defective bulbs in a box, then \$5 X is the cost of defective bulbs in a box.
a. Refer to the probability distribution shown in Exercise 5.43 and determine the probability distribution of the random variable $5X$.
b. Determine the mean daily breakdown cost, μ_{5X}, by using your answer from part (a).
c. What is the relationship between μ_{5X} and μ_X? (*Note:* From Exercise 5.43, we get μ_X.)
d. Find σ_{5X} by using your answer from part (a).
e. What is the apparent relationship between σ_{5X} and σ_X? (*Note:* From Exercise 5.43, we get σ_W.)
f. The results in parts (c) and (e) hold in general: If X is any random variable and c is a constant,

$$\mu_{cX} = c\mu_X \quad \text{and} \quad \sigma_{cX} = |c|\sigma_X.$$

Interpret these two equations in words.

5.3 The Binomial Distribution*

Many applications of probability and statistics concern the repetition of an experiment. We call each repetition a **trial,** and we are particularly interested in cases in which the experiment (each trial) has only two possible outcomes. Here are three examples.

- Testing the effectiveness of a drug: Several patients take the drug (the trials), and for each patient, the drug is either effective or not effective (the two possible outcomes).
- Weekly sales of a car salesperson: The salesperson has several customers during the week (the trials), and for each customer, the salesperson either makes a sale or does not make a sale (the two possible outcomes).
- Taste tests for colas: A number of people taste two different colas (the trials), and for each person, the preference is either for the first cola or for the second cola (the two possible outcomes).

To analyze repeated trials of an experiment that has two possible outcomes, we require knowledge of factorials, binomial coefficients, Bernoulli trials, and the binomial distribution. We begin with factorials.

Factorials

Factorials are defined as follows.

DEFINITION 5.6

What Does It Mean?

The factorial of a counting number is obtained by successively multiplying it by the next-smaller counting number until reaching 1.

Factorials

The product of the first k positive integers (counting numbers) is called ***k* factorial** and is denoted ***k*!**. In symbols,

$$k! = k(k-1)\cdots 2\cdot 1.$$

We also define $0! = 1$.

We illustrate the calculation of factorials in the next example.

EXAMPLE 5.9 Factorials

Doing the Calculations Determine 3!, 4!, and 5!.

You try it!

Exercise 5.57 on page 270

Solution Applying Definition 5.6 gives $3! = 3\cdot 2\cdot 1 = 6$, $4! = 4\cdot 3\cdot 2\cdot 1 = 24$, and $5! = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120$.

Notice that $6! = 6\cdot 5!$, $6! = 6\cdot 5\cdot 4!$, $6! = 6\cdot 5\cdot 4\cdot 3!$, and so on. In general, if $j \le k$, then $k! = k(k-1)\cdots(k-j+1)(k-j)!$.

Binomial Coefficients

You may have already encountered *binomial coefficients* in algebra when you studied the binomial expansion, the expansion of $(a+b)^n$.

DEFINITION 5.7

Binomial Coefficients

If n is a positive integer and x is a nonnegative integer less than or equal to n, then the **binomial coefficient** $\binom{n}{x}$ is defined as

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}.^{\dagger}$$

† If you have read Section 4.8, you will note that the binomial coefficient $\binom{n}{x}$ equals the number of possible combinations of x objects from a collection of n objects.

EXAMPLE 5.10 Binomial Coefficients

Doing the Calculations Determine the value of each binomial coefficient.

a. $\binom{6}{1}$ **b.** $\binom{5}{3}$ **c.** $\binom{7}{3}$ **d.** $\binom{4}{4}$

Solution We apply Definition 5.7.

a. $\binom{6}{1} = \frac{6!}{1!\,(6-1)!} = \frac{6!}{1!\;5!} = \frac{6 \cdot \cancel{5!}}{1!\;\cancel{5!}} = \frac{6}{1} = 6$

b. $\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3!\;2!} = \frac{5 \cdot 4 \cdot \cancel{3!}}{\cancel{3!}\;2!} = \frac{5 \cdot 4}{2} = 10$

c. $\binom{7}{3} = \frac{7!}{3!\,(7-3)!} = \frac{7!}{3!\;4!} = \frac{7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{3!\;\cancel{4!}} = \frac{7 \cdot 6 \cdot 5}{6} = 35$

d. $\binom{4}{4} = \frac{4!}{4!\,(4-4)!} = \frac{4!}{4!\;0!} = \frac{\cancel{4!}}{\cancel{4!}\;0!} = \frac{1}{1} = 1$

You try it!

Exercise 5.59 on page 270

Bernoulli Trials

Next we define *Bernoulli trials* and some related concepts.

DEFINITION 5.8

Bernoulli Trials

Repeated trials of an experiment are called **Bernoulli trials** if the following three conditions are satisfied:

1. The experiment (each trial) has two possible outcomes, denoted generically ***s***, for **success,** and ***f***, for **failure.**
2. The trials are independent, meaning that the outcome on one trial in no way affects the outcome on other trials.
3. The probability of a success, called the **success probability** and denoted ***p***, remains the same from trial to trial.

What Does It Mean?

Bernoulli trials are identical and independent repetitions of an experiment with two possible outcomes.

Introducing the Binomial Distribution

The **binomial distribution** is the probability distribution for the number of successes in a sequence of Bernoulli trials.

EXAMPLE 5.11 Introducing the Binomial Distribution

Mortality Mortality tables enable actuaries to obtain the probability that a person at any particular age will live a specified number of years. Insurance companies and others use such probabilities to determine life-insurance premiums, retirement pensions, and annuity payments.

According to tables provided by the National Center for Health Statistics in *Vital Statistics of the United States*, a person of age 20 years has about an 80% chance of being alive at age 65 years. Suppose three people of age 20 years are selected at random.

a. Formulate the process of observing which people are alive at age 65 as a sequence of three Bernoulli trials.
b. Obtain the possible outcomes of the three Bernoulli trials.
c. Determine the probability of each outcome in part (b).
d. Find the probability that exactly two of the three people will be alive at age 65.
e. Obtain the probability distribution of the number of people of the three who are alive at age 65.