AN INTRODUCTION TO STATISTICS

WITH

DATA ANALYSIS

by

SHELLEY RASMUSSEN

Department of Mathematical Sciences
Olney 428T
University of Massachusetts/Lowell
Lowell, MA 01854

Originally published by Brooks/Cole Publishing Company,
Division of Wadsworth, Inc.

ISBN 0-534-13578-1

Shelley_Rasmussen@uml.edu

Minitab is a statistical package, a computer program that performs many statistical procedures. The versions of Minitab now available for use on personal computers are menu-driven and much easier to use than the main-frame version originally discussed in this text. Those sections are not included in this online edition of the text. At this time, the most recent version is Minitab 15, available at very reasonable prices for purchase or rental from:

www.e-academy.com/minitab

System Requirements

Processor:	PC with a 1 GHz 32- or 64-bit processor
Memory:	512 MB or more of available RAM
Disk Space:	125 MB free space available
Operating System:	Microsoft Windows 2000, XP, or Vista.
Display:	A display capable of 1024 X 768 or higher resolution
Software:	Adobe Acrobat Reader 5.0 or higher for Meet Minitab

CHAPTER 7

Finite Probability Models Based on Counting Techniques

IN THIS CHAPTER

Permutations, combinations
Binomial distribution
Hypergeometric distribution

We mentioned in Chapter 6 that binomial experiments are commonly used by researchers in many fields. Recall that a binomial experiment consists of independent repetitions of a two-outcome experiment, with the probabilities of the two outcomes constant for each repetition. In Example 6-16, we considered a taste test of Coke versus Pepsi as a binomial experiment. Formal statistical analysis of a binomial experiment is discussed later in several sections on statistical inference (in particular, Sections 10-2, 10-5, 16-1, and 16-2). In this chapter, we develop the probability model for a binomial experiment.

In another common type of investigation, results can be summarized in a 2×2 frequency table. Our introductory example in Chapter 6 was such a study; patients were classified according to the treatment they received (new or standard) and their response to treatment (cured or not cured). One possible probability model for such an experiment leads to a probability distribution called a hypergeometric distribution. Formal statistical analysis of a 2×2 frequency table based on a hypergeometric distribution is discussed in Sections 11-5 and 16-5. We will develop the probability model that leads to a hypergeometric distribution in this chapter.

Both the binomial and the hypergeometric probability distributions involve techniques in counting. In Section 7-1, we discuss two counting problems that lead to some important distributions in statistics. One counting problem asks us to find the number of ways to arrange a finite number of objects in order (permutations). The other asks us to find the number of ways to select some of the items in a group when order of selection does not matter (combinations). After discussing permutations and combinations, we will use them to develop the binomial distribution and the hypergeometric distribution.

7-1 Permutations and Combinations

A *permutation* is an ordered arrangement of a finite number of items. How many possible ways are there to arrange n items in order? Consider first the special case in the following example.

EXAMPLE 7-1

Three United States gold medal winners in the 1988 Seoul Summer Olympics agree to be photographed for a magazine story: Theresa Edwards (basketball), Janet Evans (swimming), and Florence Griffith Joyner (track). If the three athletes stand in a row for the photograph, how many ways are there for them to line up?

The photographer has three athletes to choose from for the left position: Edwards, Evans, and Griffith Joyner. Once that athlete has been selected, the photographer has two athletes to choose from for the center spot. When this second athlete has been chosen, the third must automatically stand on the right. Therefore, there are $3 \times 2 \times 1 = 6$ possible permutations or ordered arrangements of the three women:

Edwards, Evans, Griffith Joyner
Edwards, Griffith Joyner, Evans
Evans, Edwards, Griffith Joyner
Evans, Griffith Joyner, Edwards
Griffith Joyner, Edwards, Evans
Griffith Joyner, Evans, Edwards

A shorthand notation for the product $3 \times 2 \times 1$ is 3!. We read the symbol 3! as "3 factorial."

In general, for any positive integer *n*, the symbol *n*! for *n* factorial denotes the product of all the integers from 1 to *n*. By convention, we say 0 factorial equals 1: $0! = 1$.

If we want to arrange *n* items in a row, we have *n* items to choose from for the first position. With that position filled, there are $n - 1$ remaining items to choose from for the second position, and so on. Therefore, the number of possible permutations or ordered arrangements of *n* items is the product of the integers from 1 to *n*, or *n*!.

> A **permutation** is an ordered arrangement of a finite number of objects. The number of possible permutations of *n* objects equals the product of the integers from 1 to *n*. This product is called *n* factorial and denoted *n*!.

A group of items selected from a larger collection without regard to the order of selection is a *combination*. Suppose a collection contains *n* items and *k* of these items are chosen. How many possible combinations of size *k* are there? This is the same as asking: How many ways are there to divide the *n* items into two groups, the first group with *k* items and the second group with $n - k$ items? Consider first the following special case.

EXAMPLE 7-2

Six United States gold medal winners in the 1988 Seoul Summer Olympics agree to be photographed for a magazine story: the three female athletes noted in Example 7-1, plus Carl Lewis (track), Greg Louganis (diving), and Kenny Monday (wrestling). The editor wants to divide the six athletes into two groups, one group of three to be photographed for the magazine cover, the other three to be photographed for the accompanying story. How many possible ways are there to divide the athletes into two groups of size three, a cover group and a story group?

We are not concerned here about the order of the athletes in a particular group. We care only about which athletes will be photographed for the magazine cover and which for the story inside. There are 20 ways to divide the athletes into two groups of size three, a cover group and a story group. The 20 possibilities are listed in Table 7-1.

Can we find the number 20 without writing down all 20 possible arrangements? Yes, we can. Imagine lining up the six athletes in a row and having the three athletes on the left appear in the cover photograph, the three athletes on the right in the story photograph.

TABLE 7-1 Arrangements of six athletes into two groups of size three, one group for a cover photograph and one group for a story photograph. The order within the two groups does not matter to us.

Arrangement	Cover group	Story group
1	Edwards, Evans, Griffith Joyner	Lewis, Louganis, Monday
2	Edwards, Evans, Lewis	Griffith Joyner, Louganis, Monday
3	Edwards, Evans, Louganis	Griffith Joyner, Lewis, Monday
4	Edwards, Evans, Monday	Griffith Joyner, Lewis, Louganis
5	Edwards, Griffith Joyner, Lewis	Evans, Louganis, Monday
6	Edwards, Griffith Joyner, Louganis	Evans, Lewis, Monday
7	Edwards, Griffith Joyner, Monday	Evans, Lewis, Louganis
8	Edwards, Lewis, Louganis	Evans, Griffith Joyner, Monday
9	Edwards, Lewis, Monday	Evans, Griffith Joyner, Louganis
10	Edwards, Louganis, Monday	Evans, Griffith Joyner, Lewis
11	Evans, Griffith Joyner, Lewis	Edwards, Louganis, Monday
12	Evans, Griffith Joyner, Louganis	Edwards, Lewis, Monday
13	Evans, Griffith Joyner, Monday	Edwards, Lewis, Louganis
14	Evans, Lewis, Louganis	Edwards, Griffith Joyner, Monday
15	Evans, Lewis, Monday	Edwards, Griffith Joyner, Louganis
16	Evans, Louganis, Monday	Edwards, Griffith Joyner, Lewis
17	Griffith Joyner, Lewis, Louganis	Edwards, Evans, Monday
18	Griffith Joyner, Lewis, Monday	Edwards, Evans, Louganis
19	Griffith Joyner, Louganis, Monday	Edwards, Evans, Lewis
20	Lewis, Louganis, Monday	Edwards, Evans, Griffith Joyner

There are $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ ways to line up the six athletes. But some of these arrangements will be the same as far as the cover/story division is concerned. For example, the arrangement (Edwards, Evans, Griffith Joyner, Lewis, Louganis, Monday) puts the three women on the cover and the men in the story photograph. But the arrangements (Evans, Griffith

Joyner, Edwards, Louganis, Lewis, Monday) and (Griffith Joyner, Evans, Edwards, Monday, Lewis, Louganis) do, too.

Once three athletes have been selected for the three leftmost positions, any of the $3! = 6$ permutations of these three athletes gives the same cover group. Likewise, the three athletes in the rightmost positions can be arranged in any of $3! = 6$ ways and still make up the same story group. Therefore, the $6! = 720$ permutations of the six athletes must be divided by the 3! ways to arrange the three leftmost athletes and still get the same cover group, and also divided by the 3! ways to arrange the three rightmost athletes and still get the same story group. So we find that there are

$$\frac{6!}{3!3!} = 20$$

ways to select three athletes for the cover, a photograph of the remaining three athletes going with the story. A common notation for the number of ways to divide six elements into two distinct groups of size three is $\binom{6}{3}$:

$$\binom{6}{3} = \frac{6!}{3!3!} = 20$$

In general, suppose a collection of n items is divided into two distinct groups, the first group containing k items and the second group containing $n - k$ items, where k is an integer from 0 to n. The number of ways to do this is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The symbols $\binom{n}{k}$ and $\binom{n}{n-k}$ denote the same number.

> A **combination** is a group of objects selected from a larger collection without regard to order of selection. There are
>
> $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
>
> combinations or ways to select k objects from among n objects without regard to order of selection. This is also the number of ways to divide n objects into two distinct groups, the first of size k and the second of size $n - k$.

There is exactly one way to arrange n items into two distinct groups, the first group containing n items and the second group containing no items. So, 1 must equal $\binom{n}{n}$ and $\binom{n}{0}$:

$$1 = \binom{n}{n} = \binom{n}{0} = \frac{n!}{n!0!}$$

For this reason we say 0! equals 1.

In Example 7-3, we use counting techniques to figure out how many ways a seven-game series can end.

EXAMPLE 7-3

The Atlanta Hawks and the Los Angeles Lakers are meeting in the National Basketball Association championship final series. How many possible outcomes are there in this best-of-seven game series?

Let H stand for a game won by the Hawks and L for a game won by the Lakers. Then any outcome of the series can be represented by a sequence of H's and L's. For instance, the sequence (H, L, H, H, H) denotes the outcome in which the Hawks win the championship in five games, the Lakers winning only the second game.

There are two ways the series can end in four games. These outcomes can be denoted by (H, H, H, H) and (L, L, L, L) since the series ends in four games only if one team wins the first four games.

How many ways can the series end in six games? Suppose the Hawks win the championship in six games. This means the Hawks win three of the first five games and then win the sixth, thus ending the series. (If the Hawks win four of the first five games, the series ends before the sixth game.) How many ways are there for the Hawks to win three of the first five games? The five games are divided into two groups, a group of three games won by the Hawks and a group of two games won by the Lakers. There are

$$\binom{5}{3} = \frac{5!}{3!2!} = 10$$

ways to divide the first five games into three games won by the Hawks and two games won by the Lakers. So, there are 10 ways the Hawks could win the series in six games. Similarly, there are 10 ways the Lakers could win the series in six games. Altogether there are 20 ways the best-of-seven series could end in six games, as listed in Table 7-2.

Exercise 7-6 asks you to find the number of ways the series can end in

TABLE 7-2 The number of ways a seven-game series can end in six games

Hawks win the championship	Lakers win the championship
(H, H, H, L, L, H)	(L, L, L, H, H, L)
(H, H, L, H, L, H)	(L, L, H, L, H, L)
(H, L, H, H, L, H)	(L, H, L, L, H, L)
(L, H, H, H, L, H)	(H, L, L, L, H, L)
(H, H, L, L, H, H)	(L, L, H, H, L, L)
(H, L, H, L, H, H)	(L, H, L, H, L, L)
(L, H, H, L, H, H)	(H, L, L, H, L, L)
(H, L, L, H, H, H)	(L, H, H, L, L, L)
(L, H, L, H, H, H)	(H, L, H, L, L, L)
(L, L, H, H, H, H)	(H, H, L, L, L, L)

five games and the number of ways the series can end in seven games. Putting all these totals together will give you the number of possible outcomes of such a best-of-seven series.

In Section 7-2 we use combinations to derive the binomial probability distributions. A binomial probability distribution provides the probability model for independent repetitions of a two-outcome experiment, when the probabilities of the two outcomes are the same for each repetition.

7-2 The Binomial Distributions

A coin toss is an example of an experiment with exactly two possible outcomes (heads or tails). Other examples include a taste test in which a taster must choose one of two possible products (product A or product B), an animal's survival in a toxicity study (alive or dead), a test of a product (functional or not), and a student's performance on an examination (pass or fail). We sometimes call an experiment with exactly two possible outcomes a *Bernoulli experiment* or *Bernoulli trial*.

> A **Bernoulli experiment** or **Bernoulli trial** is an experiment that has exactly two possible outcomes. For convenience, we refer to these two outcomes as success and failure.

Suppose the random variable X counts the number of successes in several independent repetitions of a Bernoulli experiment. If the probability of the success outcome is the same for each repetition, we say X has a binomial probability distribution, and we call X a binomial random variable.

In the next example, we find the probability distribution of a particular binomial random variable. Then we will discuss binomial distributions in general and look at another example.

EXAMPLE 7-4

We roll a fair six-sided die four times. (A die is *fair* if each side is equally likely to be face up at the end of a roll.) We win \$1 for each result that is divisible by 3. We win nothing otherwise. What is the probability we win at least \$2? What is our expected gain (the amount we expect to be ahead at the end of the game)? We can phrase these questions in terms of a random variable with a binomial probability distribution.

Let the random variable X equal the number of results that are divisible by 3 in the four rolls. If success refers to 3 or 6 dots coming up and failure to 1, 2, 4, or 5 dots coming up in a single roll, then X counts the number of successes in four rolls of a fair die. The probability of success equals $\frac{1}{3}$ for each roll. If the rolls are independent, then X has a binomial distribution. Let's find that binomial distribution.

Our experiment consists of four rolls of a fair die. If we assume the four rolls are independent of one another, then we can find the probability of each outcome of the experiment. The 16 possible outcomes are listed in Table 7-3, along with the value of the random variable X and the probability of each outcome.

The random variable X has five possible values: 0, 1, 2, 3, and 4. From the probabilities in Table 7-3, we can find the probability distribution for X. For example, the only way X can equal 4 is when each roll results in success (3 or 6 dots face up). Therefore, $P(X = 4) = P(S, S, S, S) = \frac{1}{81}$.

X equals 3 when the four rolls result in three successes and one failure. There are four ways the experiment could end in three successes and one failure, and each of these four outcomes has probability $\frac{2}{81}$. Therefore, $P(X = 3) = 4 \times \frac{2}{81} = \frac{8}{81}$. In a similar fashion we can find the entire probability distribution for X, shown in Table 7-4.

The probability we win at least \$2 is the probability that the random variable X is greater than or equal to 2:

$$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{24}{81} + \frac{8}{81} + \frac{1}{81} = \frac{33}{81}$$

Our expected gain is the same as the expected value of X:

$$\begin{aligned} E(X) &= \sum_{k=0}^{4} kP(X = k) \\ &= 0 \times \frac{16}{81} + 1 \times \frac{32}{81} + 2 \times \frac{24}{81} + 3 \times \frac{8}{81} + 4 \times \frac{1}{81} \\ &= \frac{108}{81} = 1\tfrac{1}{3} \text{ dollars, or about \$1.33} \end{aligned}$$

A *binomial experiment* consists of n independent repetitions of an experiment with two possible outcomes, called success and failure for convenience. The probability p of success is the same for each repetition and $0 < p < 1$. If the random variable X counts the number of successes in these n repetitions, then we say X has a binomial distribution, denoted Binomial(n, p). A binomial random variable X takes on integer values from 0 to n. Let's find $P(X = k)$ where k is an integer from 0 to n.

A **binomial experiment** consists of a finite number of independent repetitions of an experiment with two possible outcomes, called success and failure, with the probability of success the same for each repetition.

If X equals k, then k of the n repetitions of the experiment resulted in the success outcome and the other $n - k$ repetitions resulted in the failure outcome. Because the n repetitions are independent of one another, the probability of a particular sequence of k successes and $n - k$ failures is $p^k(1 - p)^{n-k}$. For example, the outcome consisting of k successes followed by $n - k$ failures has this probability. The outcome consisting of $n - k$ failures followed by k successes also has this probability. Likewise, any outcome that includes k successes and $n - k$ failures has probability $p^k(1 - p)^{n-k}$.

TABLE 7-3 The outcomes of the experiment in Example 7-4, the value of the random variable X, and the probability of each outcome

Outcome	Value of X (number of successes)	Probability
(S, S, S, S)	4	$\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{81}$
(S, S, S, F)	3	$\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{81}$
(S, S, F, S)	3	$\frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{2}{81}$
(S, F, S, S)	3	$\frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{81}$
(F, S, S, S)	3	$\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{81}$
(S, S, F, F)	2	$\frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{81}$
(S, F, S, F)	2	$\frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{81}$
(F, S, S, F)	2	$\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{81}$
(S, F, F, S)	2	$\frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{4}{81}$
(F, S, F, S)	2	$\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{4}{81}$
(F, F, S, S)	2	$\frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{4}{81}$
(F, F, F, S)	1	$\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} = \frac{8}{81}$
(F, F, S, F)	1	$\frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{8}{81}$
(F, S, F, F)	1	$\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{81}$
(S, F, F, F)	1	$\frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{8}{81}$
(F, F, F, F)	0	$\frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \frac{16}{81}$

TABLE 7-4 The binomial probability distribution for the random variable X in Example 7-4

k	$P(X = k)$
0	$1 \times \frac{16}{81} = \frac{16}{81}$
1	$4 \times \frac{8}{81} = \frac{32}{81}$
2	$6 \times \frac{4}{81} = \frac{24}{81}$
3	$4 \times \frac{2}{81} = \frac{8}{81}$
4	$1 \times \frac{1}{81} = \frac{1}{81}$

The number of ways the experiment can end in k successes and $n - k$ failures is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

This is the number of ways to divide n repetitions into two distinct groups—a group of k successes and a group of $n - k$ failures.

The event that X equals k contains the $\binom{n}{k}$ experimental outcomes corresponding to k successes and $n - k$ failures. Each of these $\binom{n}{k}$ outcomes has probability $p^k(1 - p)^{n-k}$. The probability that X equals k is the sum of these $\binom{n}{k}$ probabilities, so

$$P(X = k) = \binom{n}{k} p^k(1 - p)^{n-k}$$

for any integer k from 0 to n.

Suppose a random variable X counts the number of successes in n independent repetitions of a Bernoulli experiment, with probability p of success on each repetition. Then X has a **binomial probability distribution**, denoted **Binomial(n, p).** Probabilities for X have the form

$$P(X = k) = \binom{n}{k} p^k(1 - p)^{n-k}$$

for integer values of k from 0 to n.

We can check that this formula gives us the probabilities we found for the random variable X in Example 7-4. For instance,

$$P(X = 2) = \binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2 = \frac{4!}{2!2!} \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right)^2 = 6 \times \frac{1}{9} \times \frac{4}{9} = \frac{24}{81}$$

which is the same as the value we listed in Table 7-4.

We can find the mean and variance of a binomial random variable as follows:

If X has a Binomial(n, p) probability distribution, then the expected value of X is $n \times p$, or simply np:

$$E(X) = \sum_{k=0}^{n} kP(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k(1 - p)^{n-k} = np$$

The variance of X is $n \times p \times (1 - p)$ or $np(1 - p)$:

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{k=0}^{n} (k - E(X))^2\, P(X = k) \\ &= \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k(1 - p)^{n-k} = np(1 - p)\end{aligned}$$

The formula for the expected value of a binomial random variable makes intuitive sense. If we conduct n independent Bernoulli trials with probability p of success on each trial, then we expect, on average, the proportion p of these n trials (that is, np trials) to be successes.

In Example 7-5, we find a probability distribution for a random variable that we will see again when we discuss formal statistical analysis of a taste-test experiment in Chapter 9.

EXAMPLE 7-5

Twelve members of a statistics class participate in a taste test comparing Coke and Pepsi. Each student receives two identical-looking cups, one filled with Coke and the other with Pepsi. A student tastes samples from each cup and decides which beverage he or she prefers.

Suppose that Coke and Pepsi are equally likely to be preferred. Then how many students would we expect to express a preference for Coke? Would we be surprised if nine or more students made the same selection?

To answer these questions, we must build a probability model for our experiment. Suppose the students make independent selections, so one student's choice does not in any way affect another student's choice. Suppose also that each student has the same probability p of choosing Coke. (This second assumption may greatly simplify reality, because of physiological differences between people.) If Coke and Pepsi are equally likely to be preferred, then $p = \frac{1}{2}$.

Let the random variable Y denote the number of students who choose Coke. Then under the model assumptions above, Y has a Binomial(12, $\frac{1}{2}$) distribution. We know the probability distribution of Y without having to write down all possible outcomes in the sample space. This is fortunate; it would be extremely tedious to write down all possible sets of preferences for the 12 students, since there are $2^{12} = 4{,}096$ of them! Table 7-5 shows the probability distribution for the random variable Y.

If Coke and Pepsi are equally likely to be preferred, then we would expect half of the 12 students to express a preference for Coke and half to express a preference for Pepsi. The number of students we expect to prefer Coke is equal to the expected value of Y: $E(Y) = 12 \times \frac{1}{2} = 6$ students. The variance of Y is $\text{Var}(Y) = 12 \times \frac{1}{2} \times \frac{1}{2} = 3$ students^2 and the standard deviation of Y is $\sqrt{\text{Var}(Y)} = 1.7$ students.

Would we be surprised to see nine or more students express the same preference? The probability that nine or more students choose Coke is the probability that Y is greater than or equal to 9:

$$P(Y \geq 9) = \sum_{k=9}^{12} \binom{12}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = .073$$

TABLE 7-5 Binomial probability distribution for the random variable Y in Example 7-5. The probabilities listed have been rounded to 4 decimal places; this is why they do not sum exactly to 1.

k	$P(Y = k)$
0	$\binom{12}{0}(\frac{1}{2})^0(\frac{1}{2})^{12} = 1 \times (\frac{1}{2})^{12} = .0002$
1	$\binom{12}{1}(\frac{1}{2})^1(\frac{1}{2})^{11} = 12 \times (\frac{1}{2})^{12} = .0029$
2	$\binom{12}{2}(\frac{1}{2})^2(\frac{1}{2})^{10} = 66 \times (\frac{1}{2})^{12} = .0161$
3	$\binom{12}{3}(\frac{1}{2})^3(\frac{1}{2})^{9} = 220 \times (\frac{1}{2})^{12} = .0537$
4	$\binom{12}{4}(\frac{1}{2})^4(\frac{1}{2})^{8} = 495 \times (\frac{1}{2})^{12} = .1208$
5	$\binom{12}{5}(\frac{1}{2})^5(\frac{1}{2})^{7} = 792 \times (\frac{1}{2})^{12} = .1934$
6	$\binom{12}{6}(\frac{1}{2})^6(\frac{1}{2})^{6} = 924 \times (\frac{1}{2})^{12} = .2256$
7	$\binom{12}{7}(\frac{1}{2})^7(\frac{1}{2})^{5} = 792 \times (\frac{1}{2})^{12} = .1934$
8	$\binom{12}{8}(\frac{1}{2})^8(\frac{1}{2})^{4} = 495 \times (\frac{1}{2})^{12} = .1208$
9	$\binom{12}{9}(\frac{1}{2})^9(\frac{1}{2})^{3} = 220 \times (\frac{1}{2})^{12} = .0537$
10	$\binom{12}{10}(\frac{1}{2})^{10}(\frac{1}{2})^{2} = 66 \times (\frac{1}{2})^{12} = .0161$
11	$\binom{12}{11}(\frac{1}{2})^{11}(\frac{1}{2})^{1} = 12 \times (\frac{1}{2})^{12} = .0029$
12	$\binom{12}{12}(\frac{1}{2})^{12}(\frac{1}{2})^{0} = 1 \times (\frac{1}{2})^{12} = .0002$

Under the equal-preference model, there is about a 7% chance of seeing 9 or more of the 12 students select Coke. Similarly, under the equal-preference model, there is about a 7% chance of seeing 9 or more of the 12 students select Pepsi. So, under the equal-preference model, there is about a 14% chance that 9 or more students express the same preference. Whether we find such an event surprising depends on whether we think an event with a 14% chance of occurring to be likely or unlikely. (It would no doubt depend as well on whether we had a financial tie with one or the other beverage company!) As unbiased observers, we may not be too surprised to see 9 or more of the 12 students make the same selection, even if we believe in the equal-preference model.

Probabilities for Binomial(n, p) distributions are listed in Table A at the end of the book, for selected values of n and p.

In Section 7-3, we discuss another group of probability distributions based on combinations, the hypergeometric distributions. A hypergeometric distribution is sometimes used to model experimental results that can be summarized in a 2×2 frequency table.

7-3 The Hypergeometric Distributions

Suppose the random variable X counts the number of Type 1 objects in a sample selected at random from a finite collection of objects, each classified as either Type 1 or Type 2. Then we say X has a hypergeometric probability distribution, and we call X a hypergeometric random variable.

In the next example, we find the probability distribution of a particular hypergeometric random variable. Then we will discuss the hypergeometric distribution in general and apply it to another example.

EXAMPLE 7-6

A paper bag contains six spark plugs—four good ones and two bad ones. You are changing the spark plugs in your car, unaware that there are any bad ones in your bag. Without looking, you reach into the bag and select four spark plugs. What is the probability that at least one of the spark plugs you select is bad? What is the expected number of bad spark plugs in your sample? We can phrase these questions in terms of a random variable that has a hypergeometric probability distribution.

Let the random variable X equal the number of bad spark plugs among the four you select. If the four spark plugs in the sample were randomly selected from among the six in the bag, then X has a hypergeometric distribution. Let's find this hypergeometric distribution.

When we say that four spark plugs are randomly selected from the six in the bag, we mean that each sample of size four is equally likely to be selected. We know that the number of ways to select four items from among six items is

$$\binom{6}{4} = \frac{6!}{4!2!} = 15$$

Fifteen is the number of ways to divide the six spark plugs into two distinct groups: a group of four that is taken from the bag and a group of two that stays in the bag.

There are 15 ways to select four spark plugs from the six in the bag, and each of these outcomes is equally likely. Let's designate the four good spark plugs in the bag with the letters a, b, c, and d, and the two bad spark plugs with the letters u and v. Then we can list the 15 possible outcomes of the experiment as in Table 7-6.

The random variable X has three possible values: 0, 1, and 2. From the probabilities in Table 7-6, we can find the probability distribution for X. For instance, X equals 0 only when all four spark plugs selected are good ones, so $P(X = 0) = P(\text{a, b, c, d selected}) = \frac{1}{15}$. X equals 1 when one bad spark plug and three good ones are selected. We see in Table 7-6 that there are eight ways to select one bad spark plug and three good ones. Each of these eight outcomes has probability $\frac{1}{15}$, so $P(X = 1) = 8 \times \frac{1}{15} = \frac{8}{15}$. Similarly, we find that the probability that X equals 2 is $\frac{6}{15}$. The probability distribution for X is summarized in Table 7-7.

TABLE 7-6 The outcomes of the experiment in Example 7-6, the value of the random variable X, and the probability of each outcome. The letters a, b, c, and d represent good spark plugs; u and v represent bad spark plugs. The order within the two groups (selected and not selected) does not matter.

Outcome		Value of X, number of bad spark plugs selected	Probability
Selected	Not selected		
a, b, c, d	u, v	0	$\frac{1}{15}$
a, b, c, u	d, v	1	$\frac{1}{15}$
a, b, c, v	d, u	1	$\frac{1}{15}$
a, b, d, u	c, v	1	$\frac{1}{15}$
a, b, d, v	c, u	1	$\frac{1}{15}$
a, c, d, u	b, v	1	$\frac{1}{15}$
a, c, d, v	b, u	1	$\frac{1}{15}$
b, c, d, u	a, v	1	$\frac{1}{15}$
b, c, d, v	a, u	1	$\frac{1}{15}$
a, b, u, v	c, d	2	$\frac{1}{15}$
a, c, u, v	b, d	2	$\frac{1}{15}$
a, d, u, v	b, c	2	$\frac{1}{15}$
b, c, u, v	a, d	2	$\frac{1}{15}$
b, d, u, v	a, c	2	$\frac{1}{15}$
c, d, u, v	a, b	2	$\frac{1}{15}$

TABLE 7-7 The hypergeometric probability distribution for the random variable X in Example 7-6

k	$P(X = k)$
0	$\frac{1}{15}$
1	$\frac{8}{15}$
2	$\frac{6}{15}$

The probability that you select at least one bad spark plug is the same as the probability that the random variable X equals 1 or 2: $P(X \geq 1) = \frac{8}{15} + \frac{6}{15} = \frac{14}{15}$. We obtain the same answer by noting that the probability of at least one bad spark plug equals 1 minus the probability of no bad spark plugs selected:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{1}{15} = \frac{14}{15}$$

The expected number of bad spark plugs you select is the same as the expected value of X:

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) \\ &= 0 \times \frac{1}{15} + 1 \times \frac{8}{15} + 2 \times \frac{6}{15} = \frac{20}{15} \end{aligned}$$

or $1\frac{1}{3}$ spark plugs.

Consider now a general hypergeometric probability distribution. Suppose that in a group of N objects, m_1 are Type 1 and $m_2 = N - m_1$ are Type 2. We select a sample of size n at random from among the N objects. If the random variable X counts the number of Type 1 objects in the sample, then we say X has a hypergeometric distribution.

To specify the distribution of X, we will consider three conditions: the sample size n is less than or equal to both m_1 and m_2; the sample size n is greater than the number m_1 of Type 1 objects; the sample size n is greater than the number m_2 of Type 2 objects. (It is possible for n to be greater than both m_1 and m_2.)

Let's consider first the case that the sample size n is less than or equal to the number m_1 of Type 1 objects and the number m_2 of Type 2 objects. (This would be the case if, for example, you were selecting $n = 4$ spark plugs from a bag containing $m_1 = 5$ bad ones and $m_2 = 15$ good ones.) In this case, X can take any integer value k from 0 to n. Because of the random selection, $P(X = k)$ equals the number of ways to select k of m_1 Type 1 objects and $n - k$ of m_2 Type 2 objects, divided by the number of ways to select n of N objects.

There are $\binom{m_1}{k}$ ways to select k of m_1 Type 1 objects. There are $\binom{m_2}{n-k}$ ways to select $n - k$ of m_2 Type 2 objects. The number of ways to select k of m_1

Type 1 objects *and* $n - k$ of m_2 Type 2 objects is $\binom{m_1}{k} \times \binom{m_2}{n-k}$. The number of ways to select n objects from a total of $N = m_1 + m_2$ is $\binom{N}{n}$. Therefore, $P(X = k)$ is

$$P(X = k) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{N}{n}}$$

for integer values of k from 0 to n.

Let's consider next the second condition: the sample size n is greater than the number m_1 of Type 1 objects. (This is the case in Example 7-6, where you select $n = 4$ spark plugs from a bag containing $m_1 = 2$ bad ones and $m_2 = 4$ good ones.) Since X counts the number of Type 1 objects in the sample, X cannot be greater than m_1. The largest possible value for X is the minimum of n and m_1, denoted minimum(n, m_1). In Example 7-6, the largest possible number of bad spark plugs in the sample is 2, the minimum of $n = 4$ and $m_1 = 2$.

Now suppose the sample size n is larger than the number m_2 of Type 2 objects. Then the number of Type 1 objects in the sample is at least $n - m_2$. (This would be the case if you were selecting $n = 4$ spark plugs from a bag containing $m_1 = 5$ bad ones and $m_2 = 3$ good ones. There would have to be at least one bad spark plug in the sample, with $n - m_2 = 4 - 3 = 1$.) The smallest possible value for X in this case is the maximum of 0 and $n - m_2$, denoted by maximum(0, $n - m_2$).

Suppose a random variable X counts the number of Type 1 objects in a sample of size n selected at random from a collection of N objects, m_1 of Type 1 and $m_2 = N - m_1$ of Type 2. Then X has a **hypergeometric probability distribution,** with probabilities of the form

$$P(X = k) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{N}{n}}$$

where k is an integer from maximum(0, $n - m_2$) to minimum(n, m_1).

We can check that the formula for $P(X = k)$ works for the hypergeometric random variable in Example 7-6. You are selecting $n = 4$ spark plugs from a bag containing $m_1 = 2$ bad ones and $m_2 = 4$ good ones. The random variable X counts the number of bad spark plugs in the sample. The smallest value X can take on is maximum(0, $n - m_2$) = maximum(0, 4 − 4) = 0. The largest value X can take on is minimum(n, m_1) = minimum(4, 2) = 2. So, X can take the values 0, 1, and 2. Using the formula for hypergeometric probabilities, we find

$$P(X = 0) = \frac{\binom{2}{0}\binom{4}{4}}{\binom{6}{4}} = \frac{1 \times 1}{15} = \frac{1}{15}$$

$$P(X = 1) = \frac{\binom{2}{1}\binom{4}{3}}{\binom{6}{4}} = \frac{2 \times 4}{15} = \frac{8}{15}$$

$$P(X = 2) = \frac{\binom{2}{2}\binom{4}{2}}{\binom{6}{4}} = \frac{1 \times 6}{15} = \frac{6}{15}$$

This is the same probability distribution we found for X in Table 7-7.

We can find the mean and variance of a hypergeometric random variable as follows:

The expected value of a hypergeometric random variable X equals nm_1/N:

$$E(X) = \sum kP(X = k) = \sum k \frac{\binom{m_1}{k}\binom{m_2}{n - k}}{\binom{N}{n}} = \frac{nm_1}{N}$$

where the sum is over all integers k from maximum$(0, n - m_2)$ to minimum(n, m_1). That is, the expected number of Type 1 objects in the sample equals the sample size n times the proportion of Type 1 objects in the total collection.

The variance of a hypergeometric random variable X is

$$\begin{aligned} \text{Var}(X) &= \sum (k - E(X))^2 P(X = k) \\ &= \sum \left(k - \frac{nm_1}{N}\right)^2 \frac{\binom{m_1}{k}\binom{m_2}{n - k}}{\binom{N}{n}} \\ &= \frac{nm_1}{N} \times \left(1 - \frac{m_1}{N}\right) \times \left(\frac{N - n}{N - 1}\right) \end{aligned}$$

where the sum is again over integers k from maximum$(0, n - m_2)$ to minimum(n, m_1).

In Example 7-7, we use a hypergeometric distribution in a quality control application involving acceptance sampling.

EXAMPLE 7-7

Suppose you are in charge of quality control at Rocky's Rocking Horse Company. Each day, 90 rocking horses are produced, and you select a small number of these for careful inspection. You classify each rocking horse in your sample as either acceptable or unacceptable for shipment to toy outlets. If you find too many unacceptable rocking horses in your sample, you then inspect the entire day's production. Otherwise, you pronounce the day's production as acceptable for shipment, with no further inspection.

You know that the probability a sample passes inspection depends on the proportion of unacceptable rocking horses produced that day, as well as on the number of rocking horses you inspect. How can you use this knowledge to find an appropriate sampling and decision scheme?

Let's find out how your inspection routine will perform if you randomly select 5 rocking horses from the 90 produced; you accept the day's production lot only if all 5 are acceptable. Let the random variable W count the number of unacceptable rocking horses in your sample of 5. Then W has a hypergeometric distribution that depends on the number of unacceptable horses produced that day.

Suppose that 10% of the production is unacceptable. Then 9 of the 90 rocking horses are unacceptable and 81 are acceptable. The probability the sample passes inspection is the probability W equals 0:

$$P(W = 0) = \frac{\binom{9}{0}\binom{81}{5}}{\binom{90}{5}} = .58$$

The probability is .58 that the sample passes inspection, when 10% of the lot is unacceptable.

If 20% of the lot is unacceptable, then 18 rocking horses are unacceptable and 72 are acceptable. The probability you accept the lot is the probability W equals 0:

$$P(W = 0) = \frac{\binom{18}{0}\binom{72}{5}}{\binom{90}{5}} = .32$$

The probability is .32 that the sample passes inspection, when 20% of the lot is unacceptable.

We can continue to calculate the probability a sample passes inspection, assuming different proportions of unacceptable rocking horses in the lot. Table 7-8 lists some of these probabilities.

For a given sampling and decision rule, a plot of the probability a sample passes inspection for different proportions unacceptable in the lot is called an *operating characteristic curve.* In Example 7-7, the sampling and decision rule

TABLE 7-8 Probability of accepting the day's production lot in Example 7-7, for different proportions of unacceptable rocking horses in the lot. The decision rule is to accept the lot if no unacceptable rocking horses are found in a sample of 5 selected at random from the 90 rocking horses produced in a day.

Proportion defective in the lot	Probability of accepting the lot
$\frac{0}{90} = 0$	$\frac{\binom{0}{0}\binom{90}{5}}{\binom{90}{5}} = 1$
$\frac{9}{90} = .1$	$\frac{\binom{9}{0}\binom{81}{5}}{\binom{90}{5}} = .58$
$\frac{18}{90} = .2$	$\frac{\binom{18}{0}\binom{72}{5}}{\binom{90}{5}} = .32$
$\frac{27}{90} = .3$	$\frac{\binom{27}{0}\binom{63}{5}}{\binom{90}{5}} = .16$
$\frac{36}{90} = .4$	$\frac{\binom{36}{0}\binom{54}{5}}{\binom{90}{5}} = .07$
$\frac{45}{90} = .5$	$\frac{\binom{45}{0}\binom{45}{5}}{\binom{90}{5}} = .03$
$\frac{54}{90} = .6$	$\frac{\binom{54}{0}\binom{36}{5}}{\binom{90}{5}} = .009$
$\frac{63}{90} = .7$	$\frac{\binom{63}{0}\binom{27}{5}}{\binom{90}{5}} = .002$
$\frac{72}{90} = .8$	$\frac{\binom{72}{0}\binom{18}{5}}{\binom{90}{5}} = .0002$
$\frac{81}{90} = .9$	$\frac{\binom{81}{0}\binom{9}{5}}{\binom{90}{5}} = .000003$
$\frac{90}{90} = 1$	0

is: Select 5 rocking horses at random from among the 90 produced; accept the lot if there are no unacceptable rocking horses in the sample. Figure 7-1 shows a sketch of the operating characteristic curve for this sampling and decision rule. Some of the calculations for this curve are shown in Table 7-8. Note that the operating characteristic curve for this example is not continuous because there are just 91 possible values for the proportion of unacceptable items in the lot ($0 = \frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \ldots, \frac{89}{90}, \frac{90}{90} = 1$).

For a sampling and decision rule in an acceptance sampling inspection problem, an **operating characteristic curve** is a plot of the probability of accepting the lot for different proportions of defectives in the lot.

The probability of accepting the lot depends on both the size of the sample inspected and the decision rule. Each sampling and decision rule has its own operating characteristic curve (see Exercise 7-9).

In acceptance sampling for quality control, we compare the operating characteristic curves for several sampling and decision rules. We then select an inspection scheme that seems to best balance the opposing goals of minimizing sampling costs and accepting only lots with small proportions defective.

Note that such acceptance sampling procedures serve the *defensive* purpose of preventing unacceptable lots from being shipped to consumers (Schilling, 1982). Taking the *offensive* in quality control, we should use good experimental design during product development to design quality into the product (see Box, Hunter, and Hunter, 1978; and Part III of this book) and

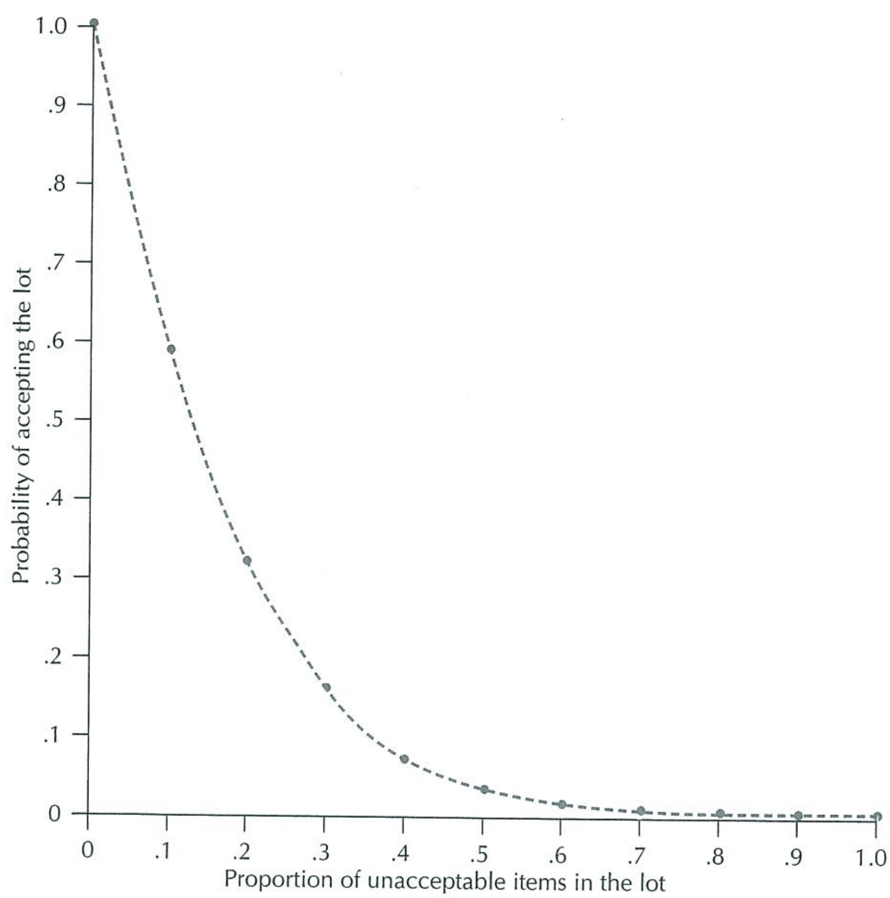

FIGURE 7-1 Operating characteristic curve for the sampling and decision rule in Example 7-7. A broken curve connects the points calculated in Table 7-8; the possible values for the proportion of unacceptable items in the lot are $0 = \frac{0}{90}, \frac{1}{90}, \frac{2}{90}, \ldots, \frac{89}{90}, \frac{90}{90} = 1$.

charting tools of statistical process control (as discussed in Ryan, 1989) to monitor quality of the product.

In Chapters 6 and 7, we have considered finite sample spaces and finite random variables. We can extend these ideas to experiments that have a countably infinite set of possible outcomes. For instance, we might count the telephone calls coming through a switchboard in a fixed interval of time, or we might count the dents in a piece of sheet metal. Then we can think of our count (or random variable) as having any nonnegative integer as a possible value. In repeated independent Bernoulli trials, if we count the number of failures before the first success, we again have a random variable that can take on any nonnegative integer value. Although interesting and useful, we will not consider probability distributions for countably infinite random variables. (See, for example, Larsen and Marx, 1986, Chapter 4; or Rice, 1988, Chapter 2.)

We will use some continuous probability distributions when we discuss statistical inference in Part III. Recall that a continuous random variable takes values in an interval of numbers. We call the probability distribution for a continuous random variable a **continuous probability distribution.** In Chapter 8, we will discuss a special group of continuous probability distributions, the Gaussian (or normal) distributions. Random variables with Gaussian probability distributions form the basis for much of classical statistical inference.

Summary of Chapter 7

Permutations and combinations form the basis for two important types of probability distributions: the binomial distributions and the hypergeometric distributions.

Suppose we have n independent repetitions of a two-outcome (success, failure) experiment, where p is the probability of success on each repetition. A random variable that counts the number of successes in these n repetitions has a binomial distribution, denoted Binomial(n, p).

Suppose a sample is selected at random from a finite collection of objects, each object classified as either Type 1 or Type 2. A random variable that counts the number of Type 1 objects in the sample has a hypergeometric probability distribution.

Minitab Appendix for Chapter 7

Finding Binomial Probabilities

Minitab provides probabilities for binomial distributions using the PDF (probability function or probability density function) command with the BINOMIAL

```
BINOMIAL WITH N =    4  P = 0.600000
   K  P( X LESS OR = K)
   0             0.0256
   1             0.1792
   2             0.5248
   3             0.8704
   4             1.0000
```

FIGURE M7-3 Cumulative probabilities for the Binomial(4, .6) distribution

Finding Cumulative Binomial Probabilities

Minitab will also provide cumulative probabilities with the CDF (cumulative distribution function) command. The command

```
MTB>   cdf;
SUBC>  binomial 4 0.6.
```

results in a display of cumulative probabilities of the form $P(X \leq c)$, where X has the Binomial(4, .6) distribution. These cumulative probabilities are displayed in Figure M7-3.

We can also have Minitab print a single cumulative probability, $P(X \leq c)$, for a specified value of c. The command

```
MTB>   cdf 3;
SUBC>  binomial 12 0.5.
      k        P(X LESS OR = k)
    3.00            0.0730
```

will cause Minitab to print the probability $P(X \leq 3)$, where X has the Binomial(12, .5) distribution. We see that $P(X \leq 3)$ equals .0730.

Finding Values Corresponding to Cumulative Binomial Probabilities

The INVCDF command with the BINOMIAL subcommand provides the number c such that $P(X \leq c)$ equals a specified probability, where X has the indicated binomial distribution. If no value of c works exactly, Minitab will print two values of c, corresponding to cumulative probabilities surrounding the specified value. For instance,

```
MTB>   invcdf 0.025;
SUBC>  binomial 12 0.5.
     k  P(X LESS OR = k)        k  P(X LESS OR = k)
     2       0.0193             3       0.0730
```

Minitab does not print probabilities for hypergeometric distributions.

Exercises for Chapter 7

EXERCISE 7-1 Four floats are lining up for the homecoming parade. How many ways are there for the four floats to line up?

EXERCISE 7-2

The starting five players of the high school basketball team are lining up for a yearbook picture. How many different ways are there to line up the players in a row? If the five players line up in a completely random fashion, what is the probability that the tallest player will be in the center position for the photograph?

EXERCISE 7-3

The final examination in your statistics class consists of eight questions. You may pick any five of the questions to answer. How many ways can you select five questions from the eight questions on the test?

EXERCISE 7-4

You and eight other students are taking a beginning statistics class. The instructor has decided that she will randomly select six of you to pass the course, and the other three to fail. How many ways are there for the instructor to choose the six students who will pass the course? What is the probability that you will pass the course?

EXERCISE 7-5

How many ways are there to select 6 elements from a set containing 36 elements, without regard to order of selection? In the Megabucks game described in Example 6-2, assume the winning numbers are selected at random. If a worthy citizen purchases one ticket with one six-number combination, what is the probability the citizen will win?

EXERCISE 7-6

Referring to Example 7-3, find the number of ways a best-of-seven game series can end in five games. Find the number of ways such a series can end in seven games. Find the total number of ways a best-of-seven game series can end.

EXERCISE 7-7

Your softball team has made it to the statewide finals. League organizers are trying to decide whether to have a five-game series or a seven-game series for the playoffs.

a. Suppose whether your team wins or loses any one game is independent of whether you win or lose any other game. What does this mean?

b. Decide whether your team has a better chance of winning a five-game series or a seven-game series if the probability you beat your opponent in any single game is: .4, .5, or .6.

EXERCISE 7-8

Suppose in Example 7-2 the editor uses a random process to choose which three of the six Olympic gold medal winners will appear on the magazine cover. Then the 20 possible outcomes in the sample space, listed in Table 7-1, are equally likely. Under this probability model, find the following probabilities:

a. The probability that the athletes on the cover will all be the same sex.

b. The probability that Florence Griffith Joyner will be on the cover. To find this probability without counting the outcomes in Table 7-1, create a hypergeometric random variable in the following way. Call Florence Griffith Joyner a Type 1 athlete and call the other five athletes Type 2. Let the random variable W count the number of Type 1 athletes on the cover. Find $P(W = 1)$.

c. The probability that Florence Griffith Joyner and Carl Lewis will be on the cover. To find this probability without counting the outcomes in Table 7-1, create a hypergeometric random variable in the following way. Call Florence Griffith Joyner and Carl Lewis Type 1 athletes and call the other four athletes Type 2. If X counts the number of Type 1 athletes in the sample, find $P(X = 2)$.

d. The probability that at least one female athlete will be in the cover photograph.

EXERCISE 7-9

You are in charge of quality control at a company that makes expensive sports cars. Factory workers produce ten cars a day. You subject all ten cars to some quality inspection. In addition, you select a sample of the ten cars for exhaustive testing. Each car tested is classified as either acceptable or not acceptable for shipment.

One day, all ten cars pass the preliminary inspection. Construct a table similar to Table 7-8 and plot the operating characteristic curve if the sampling and decision rule is:

a. Accept the day's production with no exhaustive testing.

b. Accept the ten cars produced that day if one car selected at random is found to be acceptable.

c. Accept the ten cars produced that day if two cars selected at random are both found to be acceptable.

d. Accept the ten cars produced that day if five cars selected at random are all found to be acceptable.

e. Accept the ten cars produced that day if all ten cars are found to be acceptable after exhaustive testing.

f. Compare the operating characteristic curves you found in parts (a)–(e). Discuss how each sampling and decision rule balances the opposing goals of minimizing inspection costs and accepting the day's production only when the proportion of unacceptable cars in the lot is small.

Construct a table similar to Table 7-8 and plot the operating characteristic curve if the sampling and decision rule is:

g. Accept the ten cars produced that day if, of two cars selected at random, one or none is found to be unacceptable.

h. Accept the ten cars produced that day if, among five cars selected at random, one or none is found to be unacceptable.

i. Accept the day's production if one or none of the ten cars produced is found to be unacceptable.

j. Compare the sampling and decision rules in parts (g)–(i) with each other and with the ones in parts (c)–(e).

EXERCISE 7-10 A couple plans to have two children. Suppose that they are able to carry out their plan.

a. Assume that the sex of one child is independent of the sex of the other. What does this mean?

b. Assume also that the probability p of a girl is the same for both births. Let an outcome note the sex of the firstborn child and the sex of the second-born child. Write down the sample space and probability function for this experiment.

c. If the random variable Y counts the number of female children the couple has, write down the probability distribution of Y.

d. Find the following probabilities: the probability the couple has one girl and one boy; the probability the couple has two girls; the probability the couple has two boys; the probability the couple has two children of the same sex.

e. Find the number of girls the couple can expect to have when the probability p that a baby is a girl is: $\frac{12}{25}$, $\frac{1}{2}$, or $\frac{13}{25}$.

EXERCISE 7-11 Your small cousin went trick-or-treating and has a bag of miniature candy bars. He says you can pick three without looking. You love Yummy candy bars the best. Let X denote the number of Yummy candy bars among the three you select. Find the probability distribution of X if:

a. There are five Yummy bars among ten candy bars in the bag.

b. There are two Yummy bars among ten candy bars in the bag.

c. There are eight Yummy bars among ten candy bars in the bag.

EXERCISE 7-12 An American swimmer is competing in three events in the summer Olympics. Experts give her probability p of winning, for each of the events.

a. Assume that her winning or losing one event is independent of her result in any other event. What does this mean?

b. Let an outcome note whether or not she wins the gold medal, for each of the three events. Write down the sample space and probability function for this experiment.

c. If the random variable X counts the number of gold medals the swimmer wins among the three events, write down the probability distribution of X.

d. Find the following, when the probability p she wins any single event is .4, .5, or .6: the probability the swimmer wins no gold medals; the probability she wins at least one gold medal; the probability she wins two or three gold medals; the probability she wins all three gold medals; the number of gold medals she can expect to win.

EXERCISE 7-13 At a factory there are three shifts. The numbers of men and women working each shift are summarized below:

Shift	Men	Women
Day	10	10
Evening	10	5
Night	9	3

Management announces that it will randomly select four workers from each shift for intensive performance review. When two men and two women are selected from each shift, the women claim sex discrimination. Answer the following for this problem:

a. If there really is random selection, find the probability that two or more of the four workers selected from the day shift will be women.

b. If there really is random selection, find the probability that two or more of the four workers selected from the evening shift will be women.

c. If there really is random selection, find the probability that two or more of the four workers selected from the night shift will be women.

d. If the selection is independent for different shifts, use your answers to parts (a)–(c) to find the probability under random selection that two or more women from the day shift *and* two or more women from the evening shift *and* two or more women from the night shift will be selected.

e. Use your answer to part (d) to discuss the women's claim that the selection process was *not* random, but biased toward selection of women.

f. How would your answers to parts (a)–(e) change if three women from each shift had been selected?

g. How would your answers to parts (a)–(e) change if one woman from each shift had been selected?

EXERCISE 7-14

You are a consulting statistician at a medical research center. Physicians at the center have developed a potential new treatment for a disease and wish to try it out in a preliminary study on humans. They will use the results of this study to decide whether to continue research on this treatment. The physicians come to you for advice on designing their study. To develop a probability model for their experiment, you make two simplifying assumptions:

Response to treatment is independent for different patients.
Each patient has the same probability p of being cured.

a. Discuss the reasonableness of these simplifying assumptions for this experimental situation.

If the physicians treat a total of n patients, you suggest the following decision rule. Select a number c from 0 to n. Let X denote the number of patients cured. Then decide to:

Continue research if $X \geq c$; discontinue research if $X < c$.

b. What probability distribution does X have?

c. Find $P(X \geq c)$ for the following situations:

(i) $n = 10, c = 6, p = \frac{3}{4}$

(ii) $n = 10, c = 8, p = \frac{3}{4}$

(iii) $n = 10, c = 6, p = \frac{9}{10}$

(iv) $n = 10, c = 8, p = \frac{9}{10}$

(v) $n = 20, c = 12, p = \frac{3}{4}$

(vi) $n = 20, c = 16, p = \frac{3}{4}$

(vii) $n = 20, c = 12, p = \frac{9}{10}$

(viii) $n = 20, c = 16, p = \frac{9}{10}$

d. The probability that the physicians will continue research on the treatment equals $P(X \geq c)$. Write a report to the physicians explaining how this probability depends on: the sample size n, the criterion number c selected, and the probability p of cure for any given patient. (The physicians must decide on the smallest value of p that makes the treatment worth further investigation.)

EXERCISE 7-15

Complete parts (a), (b), and (c) for each combination of sample size n and probability of success p listed in (i)–(iv):

(i) $n = 2$ and $p = \frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$

(ii) $n = 3$ and $p = \frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$

(iii) $n = 4$ and $p = \frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$

(iv) $n = 5$ and $p = \frac{1}{10}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$

a. Write down the Binomial(n, p) probability distribution.

b. Let X represent a random variable having a Binomial(n, p) distribution. Graph the probability function in a plot similar to a frequency plot (Section 2-3), with possible values k on the horizontal axis and probabilities $P(X = k)$ on the vertical axis.

c. Find the expected value and variance of each distribution from the definitions given in Section 6-8 and from the formulas given in Section 7-2.