Note to Instructor

Suggestion for introducing Chapter 5: List one frequency distribution with 10 classes having the values of *x* as 0, 1, 2, . . . , 9 and corresponding frequencies of 20, 2, 3, 2, 4, 18, 5, 4, 6, 6. List a second frequency distribution with 10 classes having the values of *x* as 0, 1, 2, . . . , 9 and corresponding frequencies of 7, 6, 6, 7, 8, 6, 7, 9, 8, 6. Inform the class that both samples are the *last digits* of recorded weights of people, but one of the samples came from *measured* weights, whereas the other sample resulted from *asking* people what they weigh. Ask the class to identify the sample with digits from reported weights. Emphasize that they can make a conclusion about the nature of the data by simply examining the *distributions.* Also ask them to construct an "ideal" distribution that would result from millions of people who were actually weighed; ask them to estimate the mean and standard deviation for this distribution. (The mean should be 4.5 and the standard deviation could be estimated using the range rule of thumb; the true mean is 4.5 and the true standard deviation is around 3.)

Here are the chapter objectives:

5-1 Probability Distributions

- Define *random variable* and *probability distribution.*
- Determine when a potential probability distribution actually satisfies the necessary requirements.
- Given a probability distribution, compute the mean and standard deviation, then use those results to determine whether results are *significantly low* or *significantly high.*

5-2 Binomial Probability Distributions

- Describe a binomial probability distribution and find probability values for a binomial distribution.
- Compute the mean and standard deviation for a binomial distribution, then use those results to determine whether results are *significantly low* or *significantly high.*

5-3 Poisson Probability Distributions

- Describe a Poisson probability distribution and find probability values for a Poisson distribution.

5-1 Probability Distributions

Key Concept This section introduces the concept of a *random variable* and the concept of a *probability distribution.* We illustrate how a *probability histogram* is a graph that visually depicts a probability distribution. We show how to find the important parameters of mean, standard deviation, and variance for a probability distribution. Most importantly, we describe how to determine whether outcomes are *significant* (significantly low or significantly high). We begin with the related concepts of *random variable* and *probability distribution.*

PART 1 Basic Concepts of a Probability Distribution

DEFINITIONS

A **random variable** is a variable (typically represented by *x*) that has a single numerical value, determined by chance, for each outcome of a procedure.

A **probability distribution** is a description that gives the probability for each value of the random variable. It is often expressed in the format of a table, formula, or graph.

In Section 1-2 we made a distinction between discrete and continuous data. Random variables may also be discrete or continuous, and the following two definitions are consistent with those given in Section 1-2.

DEFINITIONS

A **discrete random variable** has a collection of values that is finite or countable. (If there are infinitely many values, the number of values is countable if it is possible to count them individually, such as the number of tosses of a coin before getting heads.)

A **continuous random variable** has infinitely many values, and the collection of values is not countable. (That is, it is impossible to count the individual items because at least some of them are on a continuous scale, such as body temperatures.)

This chapter deals exclusively with discrete random variables, but the following chapters deal with continuous random variables.

Probability Distribution: Requirements

Every probability distribution must satisfy each of the following three requirements.

1. There is a *numerical* (not categorical) random variable x, and its number values are associated with corresponding probabilities.
2. $\Sigma P(x) = 1$ where x assumes all possible values. (The sum of all probabilities must be 1, but sums such as 0.999 or 1.001 are acceptable because they result from rounding errors.)
3. $0 \leq P(x) \leq 1$ for every individual value of the random variable x. (That is, each probability value must be between 0 and 1 inclusive.)

The second requirement comes from the simple fact that the random variable x represents all possible events in the entire sample space, so we are certain (with probability 1) that one of the events will occur. The third requirement comes from the basic principle that any probability value must be 0 or 1 or a value between 0 and 1.

EXAMPLE 1 Coin Toss

Let's consider tossing two coins, with the following random variable:

$$x = \text{number of heads when two coins are tossed}$$

The above x is a random variable because its numerical values depend on chance. With two coins tossed, the number of heads can be 0, 1, or 2, and Table 5-1 is a probability distribution because it gives the probability for each value of the random variable x and it satisfies the three requirements listed earlier:

1. The variable x is a *numerical* random variable, and its values are associated with probabilities, as in Table 5-1.
2. $\Sigma P(x) = 0.25 + 0.50 + 0.25 = 1$
3. Each value of $P(x)$ is between 0 and 1. (Specifically, 0.25 and 0.50 and 0.25 are each between 0 and 1 inclusive.)

The random variable x in Table 5-1 is a *discrete* random variable, because it has three possible values (0, 1, 2), and three is a finite number, so this satisfies the requirement of being finite or countable.

YOUR TURN Do Exercise 7 "Genetic Disorder."

TABLE 5-1 Probability Distribution for the Number of Heads in Two Coin Tosses

x: Number of Heads When Two Coins Are Tossed	$P(x)$
0	0.25
1	0.50
2	0.25