10-1 Review and Preview

A major focus of this chapter is to analyze *paired* sample data. In Section 9-4 we considered two dependent samples, with each value of one sample somehow paired with a value from the other sample. The goal in Section 9-4 was to describe methods for testing hypotheses and constructing confidence intervals for the *mean of the differences* from the matched pairs. In this chapter we again consider paired sample data, but the objective is fundamentally different from that of Section 9-4. In this chapter we introduce methods for determining whether a *correlation,* or association, between two variables exists and whether the correlation is linear. For linear correlations, we can identify an equation of a straight line that best fits the data, and we can use that equation to predict the value of one variable given the value of the other variable. We also present methods for analyzing differences between predicted values and actual values. In addition, we consider methods for identifying linear equations for correlations among three or more variables. We conclude the chapter with some basic methods for developing a mathematical model that can be used to describe some nonlinear correlations between two variables.

10-2 Correlation

Key Concept In Part 1 of this section we introduce the *linear correlation coefficient r*, which is a number that measures how well paired sample data fit a straight-line pattern when graphed. We use the sample of paired data (sometimes called **bivariate data**) to find the value of *r* (usually using technology), then we use that value to decide whether there is (or is not) a linear correlation between the two variables. In this section we consider only *linear* relationships, which means that when graphed in a scatterplot, the points approximate a straight-line pattern. In Part 2, we discuss methods for conducting a formal hypothesis test that can be used to decide whether there is a linear correlation between two variables.

Part 1: Basic Concepts of Correlation

We begin with the basic definition of *correlation,* a term commonly used in the context of an association between two variables.

DEFINITIONS

A **correlation** exists between two variables when the values of one variable are somehow associated with the values of the other variable.

A **linear correlation** exists between two variables when there is a correlation and the plotted points of paired data result in a pattern that can be approximated by a straight line.

Table 10-1, for example, includes paired sample data consisting of lengths of shoe prints and the corresponding heights of five different males. We will determine whether there is a linear correlation between the variable *x* (length of shoe print) and the variable *y* (height). Instead of blindly jumping into the calculation of the linear correlation coefficient *r,* it is wise to first *explore* the data.