## 13-5 Beyond the Basics

**13. Correcting the *H* Test Statistic for Ties** In using the Kruskal-Wallis test, there is a correction factor that should be applied whenever there are many ties: Divide $H$ by

$$1 - \frac{\Sigma T}{N^3 - N}$$

For each individual group of tied observations in the combined set of all sample data, calculate $T = t^3 - t$, where $t$ is the number of observations that are tied within the individual group. Find $t$ for each group of tied values, then compute the value of $T$ for each group, then add the $T$ values to get $\Sigma T$. The value of $N$ is the total number of observations in all samples combined. Use this procedure to find the corrected value of $H$ for Exercise 11. Does the corrected value of $H$ differ substantially from the value found in Exercise 11?

## 13-6 Rank Correlation

**Key Concept** In this section we describe the nonparametric method of the *rank correlation test,* which uses ranks of paired data to test for an association between two variables. In Chapter 10 we used paired sample data to compute values for the linear correlation coefficient $r$, but in this section we use *ranks* as the basis for computing the rank correlation coefficient $r_s$. As in Chapter 10, we should begin an analysis of paired data by exploring with a scatterplot so that we can identify any patterns in the data.

**DEFINITION** The **rank correlation test** (or **Spearman's rank correlation test**) is a nonparametric test that uses ranks of sample data consisting of matched pairs. It is used to test for an association between two variables.

We use the notation $r_s$ for the rank correlation coefficient so that we don't confuse it with the linear correlation coefficient $r$. The subscript $s$ does *not* refer to a standard deviation; it is used in honor of Charles Spearman (1863–1945), who originated the rank correlation approach. In fact, $r_s$ is often called **Spearman's rank correlation coefficient.** Key components of the rank correlation test are given in the following box, and the rank correlation procedure is summarized in Figure 13-5.

### Rank Correlation

#### Objective

Compute the rank correlation coefficient $r_s$ and use it to test for an association between two variables. The null and alternative hypotheses are as follows:

$H_0$: $\rho_s = 0$ (There is no correlation between the two variables.)

$H_1$: $\rho_s \neq 0$ (There is a correlation between the two variables.)

#### Notation

$r_s$ = rank correlation coefficient for sample paired data ($r_s$ is a sample statistic)

$\rho_s$ = rank correlation coefficient for all the population data ($\rho_s$ is a population parameter)

$n$ = number of pairs of sample data

$d$ = difference between ranks for the two values within an individual pair

*continued*