There are several formal procedures for identifying which means are different. Some of the tests, called **range tests,** allow us to identify subsets of means that are not significantly different from each other. Other tests, called **multiple comparison tests,** use pairs of means, but they make adjustments to overcome the problem of having a significance level that increases as the number of individual tests increases. There is no consensus on which test is best, but some of the more common tests are the Duncan test, Student-Newman-Keuls test (or SNK test), Tukey test (or Tukey honestly significant difference test), Scheffé test, Dunnett test, least significant difference test, and the Bonferroni test. Let's consider the Bonferroni test to see one example of a multiple comparison test. Here is the procedure.

Bonferroni Multiple Comparison Test

Step 1: Do a separate t test for each pair of samples, but make the adjustments described in the following steps.

Step 2: For an estimate of the variance σ^2 that is common to all of the involved populations, use the value of MS(error), which uses all of the available sample data. The value of MS(error) is typically obtained when conducting the analysis of variance test. Using the value of MS(error), calculate the value of the test statistic t, as shown below. The particular test statistic calculated below is based on the choice of Sample 1 and Sample 2; change the subscripts and use another pair of samples until all of the different possible pairs of samples have been tested.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\text{MS(error)} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Step 3: After calculating the value of the test statistic t for a particular pair of samples, find either the critical t value or the P-value, but make the following adjustment so that the overall significance level does not increase.

***P*-value:** Use the test statistic t with df $= N - k$, where N is the total number of sample values and k is the number of samples, and find the P-value the usual way, but adjust the P-value by multiplying it by the number of different possible pairings of two samples. (For example, with three samples, there are three different possible pairings, so adjust the P-value by multiplying it by 3.)

Critical value: When finding the critical value, adjust the significance level α by dividing it by the number of different possible pairings of two samples. (For example, with three samples, there are three different possible pairings, so adjust the significance level by dividing it by 3.)

Note that in Step 3 of the preceding Bonferroni procedure, either an individual test is conducted with a much lower significance level or the P-value is greatly increased. Rejection of equality of means therefore requires differences that are much farther apart. This adjustment in Step 3 compensates for the fact that we are doing several tests instead of only one test.

Example 2 Bonferroni Test

Example 1 in this section used analysis of variance with the sample data in Table 12-1. We concluded that there is sufficient evidence to warrant rejection of the claim of equal means. Use the Bonferroni test with a 0.05 significance level to identify which mean is different from the others.