Tukey’s HSD Test

Over the years several procedures for making multiple comparisons have been suggested. A multiple comparison procedure developed by Tukey (9) is frequently used for testing the null hypothesis that all possible pairs of treatment means are equal when the samples are all of the same size. When this test is employed we select an overall signiﬁcance level of a. The probability is a, then, that one or more of the null hypotheses is false.

Tukey’s test, which is usually referred to as the HSD (honestly signiﬁcant difference) test, makes use of a single value against which all differences are compared. This value, called the HSD, is given by

HSD = qa,k,N-k

A

MSE

(8.2.9)

n

where a is the chosen level of signiﬁcance, k is the number of means in the experiment, N is the total number of observations in the experiment, n is the number of observations in a treatment, MSE is the error or within mean square from the ANOVA table, and q is obtained by entering Appendix Table H with a, k, and N - k.

The statistic q, tabulated in Appendix Table H, is known as the studentized range statistic. It is deﬁned as the difference between the largest and smallest treatment means from an ANOVA (that is, it is the range of the treatment means) divided by the error mean square over n, the number of observations in a treatment. The studentized range is discussed in detail by Winer (10).

All possible differences between pairs of means are computed, and any difference that yields an absolute value that exceeds HSD is declared significant.

Tukey’s Test for Unequal Sample Sizes

When the samples are not all the same size, as is the case in Example 8.2.1, Tukey’s HSD test given by Equation 8.2.9 is not applicable. Tukey himself (9) and Kramer (11), however, have extended the Tukey procedure to the case where the sample sizes are different. Their procedure, which is sometimes called the Tukey-Kramer method, consists of replacing MSE/n in Equation 8.2.9 with 1MSE>2211>n + 1>n j 2, where n i and n j are the sample sizes of the two groups to be compared. If i we designate the new quantity by HSD\*, we have as the new test criterion

HSD\* = q a,k,N-k

A

MSE

2

1 1 a + b n i n j

(8.2.10)

Any absolute value of the difference between two sample means that exceeds HSD\* is declared signiﬁcant.

EXAMPLE 8.2.2

Let us illustrate the use of the HSD test with the data from Example 8.2.1.

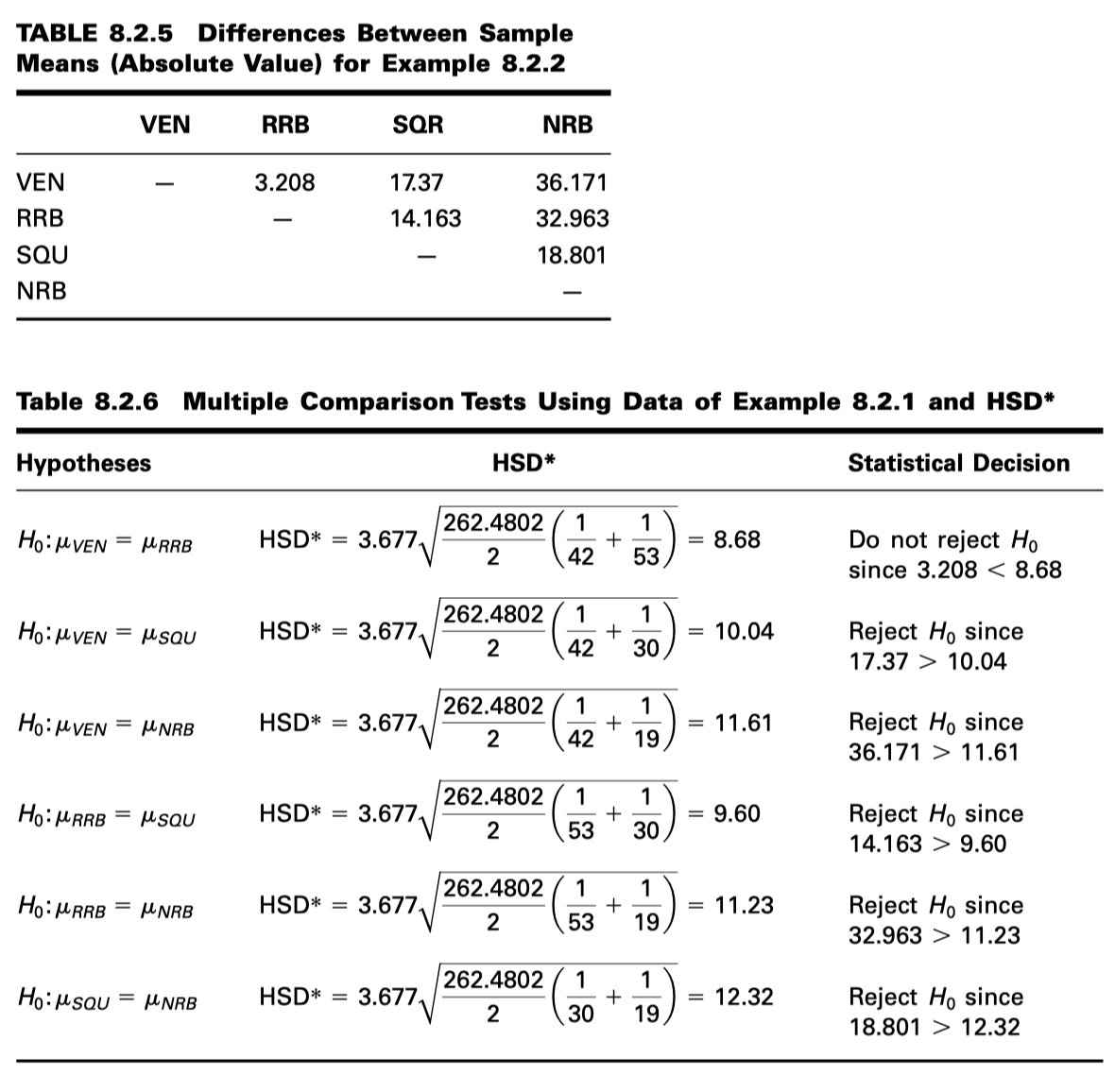
Solution:

The first step is to prepare a table of all possible (ordered) differences between means. The results of this step for the present example are displayed in Table 8.2.5.

Suppose we let a = .05. Entering Table H with a = .05, k = 4, and N - k = 140, we ﬁnd that q 6 3.68. The actual value is q = 3.667, which can be obtained from SAS® . In Table 8.2.4 we have MSE = 262.4802.

The hypotheses that can be tested, the value of HSD\*, and the statistical decision for each test are shown in Table 8.2.6.

SAS ® uses Tukey’s procedure to test the hypothesis of no difference between population means for all possible pairs of sample means. The output also contains



confidence intervals for the difference between all possible pairs of population means. This SAS output for Example 8.2.1 is displayed in Figure 8.2.8.

One may also use SPSS to perform multiple comparisons by a variety of methods, including Tukey’s. The SPSS outputs for Tukey’s HSD and Bonferroni’s method for the data for Example 8.2.1 are shown in Figures 8.2.9 and 8.2.10. The outputs contain an exhaustive comparison of sample means, along with the associated standard errors, p values, and 95% conﬁdence intervals.