Multiple Comparison Tests

If we find a significant difference using the Kruskal-Wallis test or a similar test, we may want to discover which pairs of treatments differ from each other. Since this will typically involve performing many tests, we need a procedure to adjust for the fact that we are performing many tests. The text considers two general approaches for performing multiple comparison tests, by either i) adapting traditional methods for use with ranked (or scored) data, or ii) using a permutation-based version of traditional methods. Out of the several options that are considered in the text, we will just consider three of them: i) the Bonferroni adjustment for multiple tests, ii) the rank-based version of Tukey's HSD, and iii) the permutation HSD method.

Bonferroni adjustment

Rather than using the original α level for each test, instead use $\alpha' = \alpha/m$, where m is the number of tests performed. If all pairwise tests are performed then $m = k(k-1)/2$, where k is the number of treatments.

Rank-based version of Tukey's HSD (Tukey-Kramer)

The Tukey HSD procedure is based on comparing an observed mean difference to a term proportional to the Studentized range distribution Q. This is essentially treating any mean difference as if it were selected to be between the largest and smallest observed means among the k treatments. For parametric ANOVA, the Tukey-Kramer procedure is to declare any two treatments i and j to be significantly different at level α if:

$$\left|\overline{X}_i - \overline{X}_j\right| \geq q(\alpha, k, \text{df}) \sqrt{\frac{\text{MSE}}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \ .$$

The term on the right-hand side of the inequality is called the HSD for Honest Significant Difference. The rank-based version is to declare any two treatments i and j to be significantly different at level α if:

$$\left|\overline{R}_i - \overline{R}_j\right| \geq q(\alpha, k, \infty) \sqrt{\frac{N(N+1)}{24}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \ .$$

SAS procedures, like Proc ANOVA, will give us the HSD term for the raw data, but we want the HSD for the ranked version. Note that if we multiply the top HSD term by:

$$\text{correction factor } = \frac{q(\alpha, k, \infty)}{q(\alpha, k, \text{df})}\sqrt{\frac{N(N+1)}{12 \text{ MSE}}},$$

we will then have the ranked version of the HSD. Thus we can obtain ANOVA computer output on the ranked data, multiply the HSD term from the output by the correction factor, and then we will have the correct HSD term to use for separating the ranked-data means. This offers a convenient way to use standard output to obtain the rank-based HSD procedure.

Permutation HSD

Define the statistics T_{ij} either by:

$$T_{ij} = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\text{MSE } (1/n_i + 1/n_j)}} \text{ or } T_{ij} = \frac{\overline{R}_i - \overline{R}_j}{\sqrt{(1/n_i + 1/n_j)}},$$

depending upon whether a raw-data or ranked-data based approach is desired. Once a collection of T_{ij} values is selected, a permutation HSD test can proceed by permuting the data many times as in the permutation F test, but calculating a $Q^* = \max_{ij} |T_{ij}|$ value for each permutation. The permutation distribution of Q^* can then be used for a permutation HSD procedure at described in the text. The SAS procedure MULTTEST performs these tests.