On Multiple Comparisons in R

by Frank Bretz[1], Torsten Hothorn[2] and Peter Westfall[3]

[1] Universität Hannover, Lehrgebiet Bioinformatik, Hannover

[2] Friedrich-Alexander-Universität Erlangen-Nürnberg,
Institut für Medizininformatik,Biometrie und Epidemiologie

[3] Texas Tech University, Dept. of Information Systems and Quantitative Sciences, Lubbock, TX

The `multcomp` package for the R statistical environment allows for multiple comparisons of parameters whose estimates are generally correlated, including comparisons of k groups in general linear models. The package has many common multiple comparison procedures "hard-coded", including Dunnett, Tukey, sequential pairwise contrasts, comparisons with the average, changepoint analysis, Williams', Marcus', McDermott's, and tetrad contrasts. In addition, a free input interface for the contrast matrix allows for more general comparisons.

The comparisons itself are not restricted to balanced or simple designs. Instead, the programs are designed to suit general multiple comparisons, thus allowing for covariates, nested effects, correlated means, likelihood-based estimates, and missing values. For the homoscedastic normal linear models, the program accounts for the correlations between test statistics by using the exact multivariate t-distribution. The resulting procedures are therefore more powerful than the Bonferroni and Holm methods; adjusted p-values for these methods are reported for reference. For more general models, the program accounts for correlations using the asymptotic multivariate normal distribution; examples include multiple comparisons based on rank transformations, logistic regression, GEEs, and proportional hazards models. In the asymptotic case, the user must supply the estimates, the asymptotic covariance matrix, and the contrast matrix.

Basically, the package provides two functions. The first one computes confidence intervals for the common single-step procedures (`simint`). This approach is uniformly improved by the second function (`simtest`), which utilizes logical constraints and is closely related to closed testing. However, no confidence intervals are available for the `simtest` function.

In this talk we give an introduction to the `multcomp` package. We first provide a brief theoretical background on multiple comparisons and the multiple contrast representation. We then illustrate the use of the package by going through several examples.