Nonparametric methods for clustered data in the several sample case

Erin Sprünken, Charité - Universitätsmedizin

In many trials and experiments, subjects are not only observed once, but multiple times, resulting in a cluster of possibly correlated observations. For example, mice sharing the same cage, patients providing two brain hemispheres or students of the same class are typical examples of clustered data.

Typically, under the assumption of normally distributed data, mixed models are used for analysis. However, this model assumption is rather strict and hard to justify in most real data analyses. Furthermore, skewed data (e.g. waiting times), discrete data (e.g. count data) or ordered categorical data measured on an ordinal scale are typical endpoints in a variety of trials. This motivates the use of nonparametric methods which do not rely on any specific data distribution.

For the two-sample case, several nonparametric procedures exist. For binary clustered data, a chi-square-test for contingency tables can be used. Furthermore, generalizations of the Wilcoxon-Mann-Whitney-test exist for testing the null hypothesis of equal distributions of clustered data. An extension is provided by a procedure under a less strict null hypothesis formulated in terms of the Wilcoxon-Mann-Whitney effect.

Here, we aim to generalize the procedures for the analysis of several samples. Thus, we propose a general nonparametric framework for comparing multiple groups of clustered data under mild assumptions. We present different inference methods, namely ANOVA-type test statistics and a multiple contrast test procedure and investigate their asymptotic behavior. Extensive simulation studies indicate that the methods control the type-1 error rate well, even with small sample sizes. A real data example illustrates the application of the proposed methods.