

Wilcoxon Rank-Sum Test for Clustered and Complex Survey

Stuart Lipsitz

Brigham and Women's Hospital, Harvard University

The Wilcoxon rank-sum test is one of the most frequently used statistical tests for comparing an ordered categorical outcome between two groups subjects. Even in cases where more complicated analyses are subsequently performed, initial summaries in terms of bivariate analyses are regularly reported. With a study with clustered data, such might arise in a cluster randomized study of two treatments for LDL cholesterol control ($<3.4, 3.4-4.1, 4.1-4.9, >4.9$), one would be interested in testing for equal treatment effects on LDL cholesterol control using a Wilcoxon rank-sum type test interested. However, since the subjects within a cluster are not independent, the assumptions needed for application of the Wilcoxon rank-sum test do not hold. With independent subjects, the Wilcoxon rank-sum test can be shown to be a score test from a proportional-odds cumulative logistic regression model, in which the “continuous outcome” is treated as the ordinal outcome, and the group is a dichotomous covariate; here the Wilcoxon rank-sum test is the score test for no effect of the dichotomous covariate. With clustered and complex sample survey data, we propose formulating a similar proportional-odds cumulative logistic regression model, in which the 'continuous outcome' is treated as the ordinal outcome, and the group is a dichotomous covariate. Our extension of the Wilcoxon rank-sum test to clustered and complex survey data will be a generalized estimating equations score test for no group effect under a working independence assumption for this cumulative logistic regression model. The test can easily be obtained in common statistical software such as SAS Proc Genmod or SAS Proc Survey Logistic.