Rank-Based Statistical Methods for Longitudinal Studies

Raymond G. Hoffmann
QHS (Pediatric Biostats), Pediatrics, Medical College of Wisconsin, Milwaukee
Wisconsin, 53226

Abstract

Longitudinal studies often have distributions that cannot be described by a simple generalized linear models. When this occurs an alternative distribution is needed to model the data. In particular, a rank based longitudinal method with spearman rank statistics to estimate the correlation structure will provide a more general alternative solution to this problem. It also provides a more general solution to the shape of the response curve, which might otherwise not allow comparisons between groups.

The bootstrap can be used to obtain the estimates of the functional and correlational parameters in comparison with the corresponding Wald statistics. The results would be based on simulated data that (1) can be fit by a GEE model and (2) simulated data that cannot be described by a GEE because of the shape of the response distribution.

.

Key Words: GEE ; Linear Ranks ; bootstrap

1. Linear Rank Statistics

Linear rank statistics can be applied to two types of longitudinal studies: (1) linear data models (Fitzmaurice et al and Conover et al) that is an analogue to simple repeated measures ANOVA and (2) models that are based on generalized estimating equations (GEE) or general linear mixed models (GLMM), Both assume a correlation among the repeated measures. As examples of the correlation, one can use the simplest which is compound symmetry which is a constant over all time points, a Toeplitz correlation which is a correlation matrix that is banded on both sides of the covariance, an autoregressive model which decreases as one moves away from the diagonal (c_{ii}) or an unstructured covariance model where each c_{ij} ($i<j$) can be diffeent. The two of these that are somewhat limited are the (1) autoregressive model which essentially assumes a constant time difference or integer multiple of it and (2) the unstructured model which usually has so many parameters that it does not converge.

In addition one has choose an estimate of the mean response over time. Usually the GLMM assumes a simple (first to third) order polynomial. However more general polynomials such as the fractional polynomial or more generally lowess smoothers or cubic splines give more flexibility in modeling the general shape over time. All of these require a parametric or semi-parametric model which may be the same in the two or more groups over time. Consequently, interactions and/or random effects are used to differentiate the responses in each group. .

In all or these one can use the raw data or apply a rank transform to the data to generalize the model. This is particularly important with GEE's and GLMM's which have a much more limited choice oof models; namely, a generalized linear or a member of an exponential family.

2. The Data Problem

Some data is normally distributed with a simple model for the mean and in this case. there is no need for linear rank statistics. How in many cases there are difficulties in fitting a model. For example a clinical scale, such as a pain scale, is discrete, sometimes ordinal, and may be simply skewed to the right or left. This is not a simple model and, indeed, is even more difficult if the Response is U-shaped. Which it often is. This us the situation where a linear rank transformation may provide a substantial improvement over any linear model for the data. Of course, it only works for the situation that one is comparing two models and doesn't give estimates of the shape of the response in one or more groups.

Simple simulations will use a normal underlying model to generate the discrete components of the clinical scale. Then the power of the comparison may well be substantially increased with a linear rank transform applied to the data.

Table 1: Possible Rankings of A and B Corresponding to Linear Rank Power

(R_A, R_B)	*Differences*	$\beta = .5$	$\beta = .3$	$\beta = .1$	$\beta = .01$
(1, 2,)	0	.381	.553	.819	.980
(1, 3)	1	.190	.166	.082	.010
(2, 1)	1	.190	.166	.082	.010
(2, 3)	2	.095	.050	.008	.000
(3, 1)	2	.095	.050	.008	.000
(3, 2)	3	.048	.015	.001	.000

References

Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis, 2nd edition. 2011, John Wiley.

Conover WJ and Iman RL. “Rank Transformations as a bridge between Parametric and Nonparametric Statistics. American Statistician. 1981. 124ff..

Wolfe DA. “Rank Methods”, WIRES Computational Statistics. 2009. 342-347..

McKean JW, Terpstra JT, Kioke JD. “Computational rank-based statistics” WIRES Computational Statistics. 2009. 132-140.

Hesterberg T. “Bootstrap”. WIRES Computational Statistics. 2011. 497-526..

Thompson GL. “A Unified Approach to Rank Tests for Multivariate and Repeated Measures Designs”. JASA 1991. 410-419.

Akritas MG. “The Rank Transform Method in Some Two-Factor Designs. JASA 1990. 73-78.

Blair RC, Sawilowsky SS, and Higgins JJ. “Limitations of the Rank Transformation Statistic in tests for interaction”. Communications in Statistics. 1987. 1133-1145.