

5-1-2007

JMASM 26: Hettmansperger and Mckean Linear Model Aligned Rank Test for the Single Covariate and One-Way ANCOVA Case (SAS)


Paul A. Nakonezny

The University of Texas Southwestern Medical Center

Robert D. Shull

Formerly of Feisinger Health Plan

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Nakonezny, Paul A. and Shull, Robert D. (2007) "JMASM 26: Hettmansperger and Mckean Linear Model Aligned Rank Test for the Single Covariate and One-Way ANCOVA Case (SAS)," *Journal of Modern Applied Statistical Methods*: Vol. 6: Iss. 1, Article 32. Available at: <http://digitalcommons.wayne.edu/jmasm/vol6/iss1/32>

This Algorithms and Code is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

Algorithms and Code

JMASM26: Hettmansperger and McKean Linear Model Aligned Rank Test for the Single Covariate and One-Way ANCOVA Case (SAS)

Paul A. Nakonezny

The University of Texas Southwestern Medical Center

Robert D. Shull

Formerly of Geisinger Health Plan

A SAS program (SAS 9.1.3 release, SAS Institute, Cary, N.C.) is presented to implement the Hettmansperger and McKean (1983) linear model aligned rank test (nonparametric ANCOVA) for the single covariate and one-way ANCOVA case. As part of this program, SAS code is also provided to derive the residuals from the regression of Y on X (which is step 1 in the Hettmansperger and McKean procedure) using either ordinary least squares regression (proc reg in SAS) or robust regression with MM estimation (proc robustreg in SAS).

Key words: Aligned ranks, ANCOVA, SAS, nonparametric

Introduction

Parametric analysis of covariance (ANCOVA) was introduced by Sir Ronald A. Fisher in 1932. The design goal of ANCOVA is to use the relationship between a dependent variable and covariate to adjust the dependent variable scores in order to reduce unexplained error variance (error variance in the dependent variable is reduced by an amount that can be accounted for by a covariate) and, hence, to provide a more precise estimate of treatment effects and a more powerful test of the hypothesis (Fisher, 1932; Harwell & Serlin, 1988; Maxwell & Delaney, 1990).

In order to provide a more sensitive test of the hypothesis, parametric ANCOVA must satisfy a set of underlying statistical assumptions (Elashoff, 1969; Huitema, 1980), which include (a) a linear relationship between the covariate (X) and the dependent variable (Y), (b) covariate independent of treatment, (c) equality of group

conditional variances (homogeneity of variance), (d) equality of group regression slopes (homogeneity of regression slopes), (e) normality of the distribution of Y scores for each X value within each group (conditional normality), and (f) independence of errors. However, in practical research settings, such as field experiments, it is not always possible to satisfy all of these statistical assumptions. If these underlying statistical assumptions are not tenable, then robustness and power of the parametric ANCOVA model could be threatened, and a nonparametric ANCOVA procedure should be considered.

Nonparametric ANCOVA models, which are less restrictive in their statistical assumptions, represent an alternative to the usual parametric ANCOVA. A variety of nonparametric ANCOVA models have been proposed, including those procedures developed by (1) Quade (1967); (2) McSweeney and Porter (1971); (3) Burnett and Barr (1977); (4) Rogosa (1980); (5) Conover and Iman (1982); (6) Hettmansperger and McKean (1983); and (7) Puri-Sen-Harwell-Serlin (1989). The nonparametric ANCOVA models are similar in that each model involves a ranking procedure to transform the original scores. Each model, however, is not similar with respect to robustness and power (Olejnik & Algina, 1984, 1985; Harwell & Serlin, 1988; Rheinheimer & Penfield, 2001).

Paul A. Nakonezny is Assistant Professor of Biostatistics in the Department of Clinical Sciences, Division of Biostatistics. Email: paul.nakonezny@utsouthwestern.edu. Robert D. Shull is a Quantitative Psychologist who earned a Ph.D. from the Department of Psychology at the University of Oklahoma.

Previous Monte Carlo studies that have examined the robustness and power of the nonparametric ANCOVA models have, in general, found that the Hettmansperger and McKean (1983) method (which is an aligned rank test) is robust and powerful when the underlying statistical assumptions of the parametric ANCOVA are not tenable (such as conditional non-normality, unequal regression slopes, and variance heterogeneity—even in the presence of unequal group sample sizes) (Harwell & Serlin, 1988; Rheinheimer & Penfield, 2001).

A SAS program (SAS 9.1.3 release, SAS Institute, Cary, N.C.) is presented for implementing the Hettmansperger and McKean (1983) linear model aligned rank test for the single covariate and one-way ANCOVA case. The SAS program is presented in Appendix A.

Hettmansperger and McKean (1983) Nonparametric ANCOVA Method

A brief description of the Hettmansperger and McKean (1983) nonparametric linear model ANCOVA method is presented. For a single covariate and grouping variable, the hypothesis tested based on ranks in the linear model (for the omnibus between-subjects main effect of group) is:

$$H_0: \rho_{yD_1, \dots, D_{K-1} \cdot X} = 0, \quad (1)$$

where group K and $\rho_{yD_1, \dots, D_{K-1} \cdot X}$ is the rank correlation between the dependent variable Y and the $K-1$ group membership variables D_K with the effects of the covariate X removed. Under this null hypothesis, all groups are assumed to possess identical expected covariate-adjusted mean ranks.

Hettmansperger and McKean (1983) proposed an aligned rank test (of the hypothesis specified in Equation 1) that involves the following steps: (a) calculate the least squares residuals from the regression of Y on X , where Y is the single dependent variable and X is the covariate; (b) rank the raw score residuals from low to high (1 to N); (c) weight the ranked residuals (R'_i) using $R'_i = \sqrt{12} [R_i / (N + 1) - 0.5]$, where R_i is the ranked

raw score residuals, N is the total number of observations, 12 and 0.5 are constants; and (d) perform a parametric ANOVA (with least squares estimation) on the weighted ranked residuals (R'_i), while maintaining the original group membership.

The sum of squares between groups, SS_{Between} , from the parametric ANOVA on R'_i is the aligned rank *Test Statistic*, which is asymptotically distributed as a central χ^2 with $K-1$ degrees of freedom. Hettmansperger and McKean (1983) originally proposed the use of least squares residuals from the regression of Y on X (step 1 in their procedure). If outliers are present in the data, however, then the residuals from the regression of Y on X (step 1 in the procedure) should be derived using robust regression. The current article, therefore, provides SAS code to derive the residuals using either ordinary least squares regression (proc reg in SAS) or robust regression with MM estimation (proc robustreg in SAS). The reader is referred to Rousseeuw and Leroy (1987) and Yohai (1987) for a discussion of robust estimates for regression.

Although the SAS program in the current article only addressed the single covariate and one-way ANCOVA case (for the one-tailed test of the omnibus between-subjects main effect of group), the Hettmansperger and McKean (1983) aligned rank test can be extended to the multiple covariate case and two-way/higher-order factorial ANCOVA case. A separate aligned rank test, however, is required for each main effect and interaction effect tested. The reader is referred to Adichie (1978), Hettmansperger and McKean (1983), and Hettmansperger (1984) for an expanded discussion of the aligned rank test.

References

- Adichie, J. N. (1978). Rank tests of subhypotheses in the general linear regression. *Annals of Statistics*, 6, 1021-1026.
- Burnett, T. D., & Barr, D. R. (1977). A nonparametric analogy of analysis of covariance. *Educational and Psychological Measurement*, 37, 341-348.

- Conover, W. J., & Iman, R. L. (1982). Analysis of covariance using the rank transformation. *Biometrics*, 38, 715-724.
- Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 6, 383-401.
- Fisher, R. A. (1932). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.
- Harwell, M. R., & Serlin, R. C. (1988). An empirical study of a proposed test of nonparametric analysis of covariance. *Psychological Bulletin*, 104, 267-280.
- Harwell, M. R., & Serlin, R. C. (1989). A nonparametric test statistic for the general linear model. *Journal of Educational Statistics*, 14, 351-371.
- Hettmansperger, T. P., & McKean J. W. (1983). A geometric interpretation of inferences based on ranks in the linear model. *Journal of the American Statistical Association*, 78, 885-893.
- Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.
- Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data*. Belmont, CA: Wadsworth Publishing Company.
- McSweeney, M., & Porter, A. C. (1971, April). *Small sample properties of nonparametric index of response and rank analysis of covariance*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Olejnik, S. F., & Algina, J. (1984). Parametric ANCOVA and the rank transform ANCOVA when the data are conditionally non-normal and heteroscedastic. *Journal of Educational Statistics*, 9, 129-149.
- Olejnik, S. F., & Algina, J. (1985). A review of nonparametric alternatives to analysis of covariance. *Evaluation Review*, 9, 51-83.
- Puri, M. L., & Sen, P. K. (1969). Analysis of covariance based on general rank scores. *Annals of Mathematical Statistics*, 40, 610-618.
- Quade, D. (1967). Rank analysis of covariance. *Journal of the American Statistical Association*, 62, 1187-1200.
- Rousseeuw, P.J. & Leroy, A.M. (1987). *Robust regression and outlier detection*. New York: John Wiley & Sons, Inc.
- Rheinheimer, D. C., & Penfield, D. A. (2001). The effects of Type I error rates and power on the ANCOVA F test and selected alternatives under nonnormality and variance heterogeneity. *The Journal of Experimental Education*, 69, 373-391.
- Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin*, 88, 307-321.
- SAS. Cary, NC: Statistical Analysis System Institute.
- Yohai, V. J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression. *Annals of Statistics*, 15, 642-656.

Appendix A

The following SAS program (SAS 9.1.3 release) can be used to perform the Hettmansperger and McKean (1983) aligned rank test for the single covariate and one-way ANCOVA case (one-tailed test of the omnibus between-subjects main effect of group).

```
/* Step 1: Read internal data into SAS data set hett */
```

```
/* Note. You may also read data from an external file into a SAS data set using Proc Import */
```

```
data hett;
input subject group X Y;
cards;
<your data>
;
```

```
/* Step 2: Use Proc Reg to compute and output the least squares residuals from the regression of Y on X, where Y is the dependent variable and X is the single covariate */
```

```
proc reg;
  model Y=X;
  output out=res r=residual;
run;
proc print data=res;
```

```
/* Alternative Step 2: If outliers are present in the data, then the residuals should be derived using robust regression. Use Proc Robustreg to compute and output the robust residuals, with MM estimation, from the regression of Y on X */
```

```
proc robustreg method=mm;
  model Y=X;
  output out=res r=residual;
run;
proc print data=res;
```

```
/* Step 3: Use Proc Rank to rank the raw score residuals from low to high (1 to N) */
```

```
proc rank data=res out=rank; var residual;
```

```
/* Step 4: Weight the ranked residuals ( $R'_i$ ) using  $R'_i = \sqrt{12} [R_i / (N + 1) - 0.5]$ , where  $R_i$  is the ranked raw score residuals and N is the total number of observations */
```

```
data weight; set rank nobs=n;
  rename residual=RankedResiduals;
  weight=12** .5*(residual/(n+1)-.5);
run;
proc print data=weight;
```

```
/* Step 5: Perform a parametric ANOVA on the weighted ranked residuals ( $R'_i$ ), while
maintaining the original group membership */

proc glm; class group;
  model weight=group;
  ods output modelanova=ma(where=(hypothesistype=3));
run;

/* Step 6: One-tailed test of the omnibus between-subjects main effect of group
  Dependent = dependent variable = weighted ranked residuals ( $R'_i$ )
  Source = omnibus main effect of group
  DF = K-1 degrees of freedom
  HettMcKeanChisqValue = aligned rank test statistic (which is the  $SS_{\text{Between}}$ )
  ProbChisq = p-value of the obtained test statistic
  CriticalValueChisq = chi-square critical value with K-1 degrees of freedom,  $\alpha = .05$  */

data ma;
  set ma;
  rename ss=HettMcKeanChisqValue;
  ProbChisq=sdf('chisquare',ss,df);
  CriticalValueChisq=quantile('chisquare',.95,df);
run;

title 'The Hettmansperger and McKean Nonparametric One-Way ANCOVA (aligned rank test)
Procedure';
proc print data=ma noobs; var Dependent Source DF HettMcKeanChisqValue ProbChisq
CriticalValueChisq;
run;
```