

11-1-2002

Type I Error Rates For Rank-Based Tests Of Homogeneity Of Slopes

Alan J. Klockars
University of Washington

Tim P. Moses
University of Washington

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

Recommended Citation

Klockars, Alan J. and Moses, Tim P. (2002) "Type I Error Rates For Rank-Based Tests Of Homogeneity Of Slopes," *Journal of Modern Applied Statistical Methods*: Vol. 1 : Iss. 2 , Article 54.
DOI: 10.22237/jmasm/1036109520

Type I Error Rates For Rank-Based Tests Of Homogeneity Of Slopes

Alan J. Klockars
Educational Psychology
University of Washington

Tim P. Moses
Educational Psychology
University of Washington

The purpose of this study was to explicate two issues concerning the standard and rank based test of homogeneity of slopes. Two alternative ranking methods intended to address nonnormality and additive treatment effect patterns were developed and compared in terms of their ability to control Type I error. The results replicated previous findings of inflated Type I error rates with leptokurtic curves and with rank based tests with some patterns of additive treatment effects. The new nonparametric procedures generally control Type I error although they were slightly inflated with skewed distributions.

Key words: Slope homogeneity, ranking methodology, type I error

Introduction

Psychology and education have long acknowledged the need for methods to address the interaction between treatment variables on the one hand and individual difference variables on the other. Cronbach (1957) in his presidential address to the American Psychological Association called for a fusion of the “two schools of psychology”, a field later to be identified as Aptitude x Trait interaction (ATI) research (Cronbach & Snow, 1981). While ATI research was originally developed within educational psychology it has spread throughout psychology including industrial psychology (see for instance, Hunter, Schmitt & Hunter, 1979) and psychotherapy (see for instance

Dance & Neufeld, 1988). Two major strategies are used to explore ATIs. The first is based on stratification of the individual difference variable, which produces a randomized block design. The desired information is contained in the Block x Treatment interaction. The alternative is a regression based approach that can be viewed either as a test of moderated regression or of homogeneity of slopes within an analysis of covariance design.

The usual form of the regression approach is to assume a linear relationship between the individual difference variable used as the covariate (X) and the outcome measure (Y). The issue investigated is whether the treatment alters the nature of the linear relationship. The presence of an interaction between the treatment and X is reflected in the difference between the slopes. This finding may be the primary finding of the study and may also inform the researcher regarding appropriate strategies for looking for main effects.

We will adopt the regression vantage point for describing the issues addressed. Cronbach and Snow (1981) argued for the regression approach as more powerful than stratification, an assertion that was supported in simulations by Klockars and Beretvas (2001). The issue of power is particularly important given the high Type II error rates associated with attempts to identify interactions, especially in field studies (McClelland & Judd, 1993). For a comparison of randomized block and analysis of covariance see

Alan J. Klockars is Professor of Educational Psychology at the University of Washington. His research concerns multiple comparisons and, more recently, methods of conducting ATI research. Address correspondence to Alan J. Klockars, Box 353600, University of Washington, Seattle, WA 98195-3600. E-mail: klockars@u.washington.edu. Tim P. Moses is a doctoral candidate in the Educational Psychology program at the University of Washington. His research and teaching focus on the application of statistics to the study of social phenomena and the influences of assumption violations on the accuracy of standard and alternative statistical methods.

Klockars, Potter, and Beretvas (1999), and Klockars and Beretvas (2001).

The test of homogeneity of slopes is based on a set of assumptions common to both regression and covariance. Of primary importance in the current investigation is the assumption that the variables are normally distributed. The assumption is part of a mathematical model and, as with any model, it is unexpected that empirical data will ever exactly fulfill the model (e.g. scores are discrete while the model is continuous).

However, Micceri (1989) in a survey of typical variables analyzed in psychology and education journals reported that the distributions were often far from normal with considerable skew and kurtosis. Conover and Iman (1982) and more recent work by Headrick and Sawilowsky (2000) showed that the Type I error control of the test of homogeneity of slopes is greatly impacted by the shape of the distributions involved. Platykurtic or light-tailed distributions produce Type I error rates that are conservative while leptokurtic or heavy-tailed distributions produce liberal Type I error rates. Klockars and Moses (2001) found that the Type I error rates for distributions with shapes that Micceri (1989) indicated were typical far exceeded both Bradley's (1978) conservative (.055) and liberal (.075) definition of robustness.

Prior research has not directly addressed the question of the relative impact of nonnormality in X and Y on Type I error. Atiquallah (1964) showed analytically that the shape of the distribution of X plays a role in the magnitude of the calculated F ratio as does the distribution of Y. In simulation studies three different patterns of X and Y distributions have been used.

Conover and Iman (1982) and Stephenson and Jacobson (1988) varied the shape of the Y distribution but used a normally distributed X distribution throughout. Headrick and Sawilowsky (2000) let the X and Y distributions have the same shape so that if Y were moderately right skewed the X distribution would also be moderately right skewed. Klockars and Moses (2001) systematically varied the shape of the Y distribution and created the X distribution as a linear combination of Y and normally distributed random error. Thus the covariate, X, had a distribution less extreme than that of the Y distribution. This was particularly true with the

low correlation condition in which the normally distributed random error was more heavily weighted.

The first issues under investigation in the current study are (1) a replication of the finding that the shape of the Y distribution systematically influences Type I error rates of the test of homogeneity of slopes, and (2) an evaluation of the relative importance and independence of the shape of the X distribution compared to that of the Y distributions in producing Type I errors.

A number of authors have proposed non-parametric, rank based analyses of covariance to avoid the distributional requirements of analysis of covariance as a test of adjusted means (Quade, 1967; Puri & Sen, 1969; Burnett & Barr, 1977; Shirley, 1981). These strategies, however, focused primarily on the null hypothesis regarding the adjusted means of the treatment groups. Slopes were assumed to be homogeneous and the question of an interaction was not addressed.

Shirley (1981) developed χ^2 tests for both the test of parallel lines and equal adjusted means on data where the outcome measure Y was converted to ranks. Conover and Iman (1982) proposed standard analysis of covariance on data where both X and Y were replaced with their ranks. Stevenson and Jacobsen (1988) offered a "hybrid" alternative in which only the Y variable was ranked while X was retained in its original form. A standard ANCOVA was conducted on the raw X and ranked Y scores to test for both differences in slopes and adjusted means. In the latter two studies simulated data were generated to evaluate how robust the methods were. The rank and hybrid ANCOVA methods tended to control Type I error in situations where the error rate for the original observations was problematic, that is, where the Y distributions were leptokurtic.

More recent inquiries using analysis of covariance with ranks have returned to considering only questions about the adjusted means (Seaman, Algina, & Olejnik, 1985; Harwell & Serlin, 1988; Hettermansperger, 1984; Rheinheimer & Penfield, 2001). However, Headrick and Sawilowsky (2000) presented simulation evidence that indicated that the Conover and Iman approach to testing differences in slopes can have very elevated Type I error rates under conditions of additive treatment effects. In particular, simulations in which a small proportion of the treatment effects had large

additive effects resulted in extremely high Type I error rates when the test for homogeneity of slopes was conducted. When X and Y were highly correlated and the sample size was large there was essentially a 100% chance of rejecting the null hypothesis that the slopes differed. This happened even though the only effects built into the data were additive effects that should have been reflected in the test of adjusted means rather than slopes. The present study (3) replicates the Headrick and Sawilowsky finding and (4) develops alternative methods for testing for differences in slopes within the general analysis of covariance framework that may have better control of Type I error.

The development of alternative non-parametric methods relies on understanding why there is an elevated level of Type I error when additive treatment effects are present. Let the parameters of the original measurements be indicated by standard Greek letters with X, Y, and k subscripts, and those of the ranked scores by Greek letters with the addition of the subscript R to denote ranked. The null hypothesis in a test of homogeneity of slopes for the original scores is $\beta_1 = \beta_2 = \dots = \beta_k$ with each of the slopes given by

$$\beta_k = \rho_k \frac{\sigma_{Y_k}}{\sigma_{X_k}} \tag{1}$$

We dealt with the case where the null hypothesis concerning slopes implies equality of the elements on the right side of equation 1. If the null hypothesis for slopes is true then the variability of the X scores, the Y scores and the XY correlations are homogeneous. The special case where the slopes are equal because of compensating effects such as inversely related correlations and Y variances was not considered.

The question of interest concerns the equality of the β_k s but is tested by evaluating the null hypothesis concerning the equality of the β_{Rk} s. This will be an equivalent test if the terms on the right side of:

$$\beta_{Rk} = \rho_{Rk} \frac{\sigma_{RY_k}}{\sigma_{RX_k}} \tag{2}$$

are homogeneous when the terms on the right side of (1) are homogeneous.

The variances of the raw X scores (σ^2_{Xk}) are homogeneous by the nature of an experiment. Under the standard procedures associated with ANCOVA the subjects are randomly assigned to conditions with no impact of treatment present in the X scores. The variances for the ranked X scores, σ^2_{RXk} , will be $((kn)^2 - 1)/12$ and the sampled set of ranks from all k groups should estimate this parameter because of the random assignment. Additive treatment effects will have no impact on either σ^2_{Xk} or σ^2_{RXk} .

The correlation between the ranked XY scores (ρ_{Rk}) will be similar but not identical to the correlation between the original scores. If the treatment conditions have equal correlations in their raw score form, that equality of correlation will be maintained in the ranked scores. Additive treatment effects should have no or only minor influences on the homogeneity of correlations based on ranked scores.

As with the ranked X scores, the variance of the ranked Y scores is a simple function of sample size (n) and number of groups (k). If there are no additive treatment effects the variance of the ranked Y scores is:

$$\sigma^2_{RY_k} = \frac{(kn)^2 - 1}{12} \tag{3}$$

Additive treatment effects have the possibility of changing the variance of ranked Y scores within a group. Since group slopes are a function of the standard deviations of X and Y, additive treatments could produce the appearance of an interaction. This possibility is most easily seen in an exaggerated example. Consider the pattern of treatment effects for 4 groups of {0, 0, 0, c} where c is an additive constant. Let c be so large that the fourth sample of scores is raised so that no member of group 4 has a score lower than the highest score in the remaining groups. In this case the ranked Y variances estimated by the first 3 of the k=4 groups would be:

$$\sigma^2_{RY_k} = \frac{((k-1)n)^2 - 1}{12} \tag{4}$$

while the variance of the fourth group would reflect the variability in ranks of n adjacent scores which is

$$\sigma^2_{RY_4} = \frac{n^2 - 1}{12} \quad (5)$$

The differences in the variability from equation 4 and 5 would produce a set of slopes in which the last group would have a slope almost $k-1$ times smaller than the slopes of the remaining groups. For smaller additive treatment effects the separation would be less complete but still result in the reduction of the Y variability for the separated group and thus a reduction in the slope. Headrick and Sawilowsky's (2000) report of high rejection rates of the null hypothesis concerning equal slopes are Type I errors in the sense that there were only additive rather than interactive effects present. The rejections are also correct rejections of the null hypothesis concerning slopes after the additive treatment effects have confounded additive and interactive effects when Y is ranked. The proportion of rejected hypotheses will depend on the power, which is a function of the correlation and sample size.

Other configurations of additive effect would not produce the same effect. Patterns such as $\{0, 0, c, c\}$ or $\{-2c, -c, 0, c, 2c\}$ would alter all of the groups' Y variabilities equally and thus retain equal slopes in the ranked scores if there were equal slopes in the original distributions. In the simulations performed by Stephenson and Jacobson (1988) the vector of additive effects was $(1, 0, 1.5, 3)$. This pattern did not produce inflated Type I error rates as the spacing is relatively equal and the sample size and correlation were much lower than in Headrick and Sawilowsky, providing little power.

To eliminate the potential of additive treatment effects confounding the test of differences in slope we propose that the ranking of observations be based on a function of the scores that would eliminate any additive effects. The first alternative is to subtract the appropriate group sample mean from each score prior to ranking the observations and conducting the analysis of covariance. Scores within a treatment condition are defined as $Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$. The sample mean has an expected value of $\mu + \alpha_j$. Analysis of the

deviation from the group mean provides estimates of a common ε_{ij} .

The second alternative is to subtract the sample median prior to ranking the observations. Like the sample mean, the sample median will cancel additive treatment effects. Any constant difference reflecting the difference between the population mean and median should be eliminated when the differences are ranked. The median is offered as an alternative when the distribution of Y scores may be highly skewed.

Consider the situation in which the null hypothesis concerning slopes is true but the outcome measure is a right skewed, heavy-tailed distribution. The presence in a sample of a single, outlying score would produce deviations from the mean that were primarily negative, reflecting the inflating effect of the extreme score on the sample mean. The predominance of negative deviations along with the outlying positive deviation would distort the slope and inflate the Type I error rate.

A number of robust statistics are available to decrease the influence of extreme scores. The sample median is one of the simplest and is used in the current alternative approach. In both proposed methods the test for additive treatment effects would have to be conducted using the normal Conover and Iman (1982) or alternative method. The subtraction of either the sample mean or the sample median from scores eliminates any additive effects and precludes the deviations from being used to evaluate additive effects.

Methodology

All simulations were conducted on a Unix computer using programs written in FORTRAN 77. Unit normal distributions were generated using the RNNOR subroutine of IMSL. All Type I error rates were obtained from 50,000 iterations of the program. For the nominal value of .05 this number of iterations produces a standard error of .001. The simulations were all based on a one-way design with $k=4$ groups, $n=20$ subjects, and a single covariate. Two levels of relationship between X and Y were created to represent a relatively low and relatively high degree of relationship. In the normally distributed X and Y scores the two levels represent correlations of .3 and .7.

The normally distributed covariate X was generated by RNNOR. The Y variable was created as a weighted linear combination of X and a second randomly created normal distribution to introduce random error. The weights were selected so that the variance of the Y scores was 1 and the slopes for all groups would be either .3 or .7. The original normally distributed X and Y variables (NOR X and NOR Y, respectively) were then transformed to three other shaped distributions using Fleishman's (1978) power vector method.

A platykurtic distribution was selected for study with skew of 0 and kurtosis of -1 (PLAT X and PLAT Y). The other two distribution were leptokurtic, the first with skew of 0 and kurtosis of 1.5 (LEPTO X and LEPTO Y) and finally, a more extreme, skewed, leptokurtic distribution with skew of 1.75 and kurtosis of 3.75 (SKLPT X and SKLPT Y).

All 16 possible combinations of shape of X and shape of Y were analyzed. Because of the multiple pairings no attempt was made to correct the correlations to exactly .3 and .7 in all pairings (see Headrick and Sawilowsky, 1999). The actual correlations for the 16 pairings varied from .22 to .30 for the nominal .3 and from .55 to .70 for .7. The first three shapes with no skew had much more homogeneous correlations, ranging from .28 to .30 and from .66 to .70 for .3 and .7, respectively. We shall refer to the two conditions as Low and High correlation, respectively.

Three configurations of additive treatment effects were used to evaluate the previously reported confounding of additive treatment effects with the test of slopes. The first condition had no additive effects. The second and third had configurations of 0, 0, 0, c and 0, 0, c, c, respectively. The four levels of additive constant c were .8, 1.4, 2.0, and 2.6. This produced $1+(2)(4)=9$ distinct patterns. Because both X and Y have unit variance the additive constants are in z-scores.

Each data set was analyzed with four representations of the data. These are:

- | | |
|-----------------------|---|
| 1. X- Original Scores | Y-Original Score (XY) |
| 2. X- Ranked Scores | Y-Ranked Scores (RxRy) |
| 3. X- Ranked Scores | Y-Ranked deviation from sample mean (RxR1y) |
| 4. X- Ranked Scores | Y- Ranked deviation from sample median(RxR2y) |

The analysis of the data set (XY) is the standard parametric analysis of covariance, the second (RxRy) is the Conover and Iman (1982) non-parametric analysis of covariance, the third (RxR1y) and fourth (RxR2y) are the non-parametric analyses of covariance developed in the current paper based on the mean and median, respectively.

Results

The results were obtained by averaging the probabilities of Type I error across the simulations. The primary findings are a description of those variables that impact Type I error. In addition each Type I error rate is classified as to whether it exceeds either Bradley's (1978) conservative or liberal criterion for robustness. Although these criteria are arbitrary they provide a commonly known standard for evaluating the magnitude of the elevation of Type I error.

The first two issues deal with the relationship between the shape of the underlying distribution and Type I error. The analyses are based on the conventional analysis of covariance of the original scores, XY. Table 1 contains the mean Type I error rates for all combinations of shapes for X and Y. The results are presented separately for the low and high correlation conditions. Each mean is based on the simulations representing the nine different additive treatment combinations. Preliminary analyses indicate that additive configurations and magnitude represent trivial factors and could be combined without loss of information. Also included in Table 1 are the number of the simulations that had Type I error rates that exceeded Bradley's conservative (.055) and liberal (.075) criterion level for robustness. Only the upper limits are considered, as the present concern is for unacceptably high Type I error rates. Low error rates are more likely to be reflected in poor power.

The average Type I error rate for both LEPTO Y and SKLPT Y are considerably larger than for the normal curve. PLATY has a conservative Type I error rate. The inflated Type I error rates associated with leptokurtic curves is further seen in the frequency with which the Type I error rate exceeds even the most liberal of robustness criteria.

Table 1. Average Type I error rates across raw X and Y distributions and correlations.

Corr.	Y Dist.	X Distribution			
		PLAT X	NOR X	LEPTO X	SKLPT X
LOW $r \approx .3$	PLAT Y	.044 (0,0)	.041 (0,0)	.040 (0,0)	.043 (0,0)
	NOR Y	.050 (0,0)	.050 (0,0)	.051 (0,0)	.052 (0,0)
	LEPTO Y	.056 (7,0)	.059 (9,0)	.062 (9,0)	.062 (9,0)
	SKLPT Y	.060 (9,0)	.068 (9,0)	.072 (9,0)	.124 (9,9)
HIGH $r \approx .7$	PLAT Y	.025 (0,0)	.020 (0,0)	.034 (0,0)	.051 (0,0)
	NOR Y	.055 (6,0)	.049 (0,0)	.058 (9,0)	.068 (9,0)
	LEPTO Y	.083 (9,9)	.094 (9,9)	.103 (9,9)	.094 (9,9)
	SKLPT Y	.114 (9,9)	.178 (9,9)	.213 (9,9)	.225 (9,9)

Note. Numbers in parentheses are the number of times Type I error exceeded .055 and .075 in that condition where the maximum is 9.

The variability in the means presented in Table 1 is partitioned into the main effects and interactions between the independent variables in the simulation. Table 2 contains the mean square deviations for these sources. Because of the number of iterations all of the effects are significant based on the most conservative of standards. In the current discussion it is the relative size of the effects that is of primary concern.

Three effects are much larger than the remaining sources. These are the shape of the original Y distribution, the strength of the XY correlation, and the interaction between the shape of the Y distribution and the correlation. The shape of the X distribution has a mean square less than one-tenth that of the shape of the Y distribution. The interaction between the shapes of X and Y is small and trivial.

Table 2. Sources of variation on Type I error rate with raw X and Y scores (XY).

Source	SS	df	MS
Y Distribution shapes	0.365	3	0.122
Correlation (COR)	0.079	1	0.079
Y * COR	0.129	3	0.043
X Distribution shapes	0.034	3	0.011
X*Y	0.044	9	0.005
X * COR	0.008	3	0.003
X*Y * COR	0.012	9	0.001
Residual	0.0004	256	0.000

The main effect for the shape of Y reflects the variability in the overall means. The interaction is reflected in Type I error rates that are more extreme with a higher correlation. Platykurtic curves become more conservative and leptokurtic curves more liberal. Because there were more leptokurtic curves than platykurtic curves the average Type I error rate for the higher correlation is larger. The pattern of the means for the shapes of the X distribution mirror those of Y but are much less extreme.

The next two issues deal with the ability of ranking methods to control Type I errors for differences in slopes when there are additive treatments present. Table 3 presents the Type I error rates for the ANCOVA test of slopes proposed by Conover and Iman (1982), the ANCOVA test of slopes based on deviations of scores from the appropriate sample mean, and the ANCOVA test of slopes based on deviations of scores from the appropriate sample median. Two patterns of treatment effect, {0,0,0,c} and {0, 0, c, c} are paired with four levels of c. The results are summed across the 4x4=16 distributional pairings. Results are reported separately for low and high correlations. The parenthetical values indicate how many of these 16 simulations produced Type I error rates that exceeded the conservative and liberal robustness criteria.

Table 3. Average Type I error across correlation, treatment effect, treatment effect pattern and ranking method.

Corr.	Pattern	Data Set	Treatment Effect (c)			
			.8	1.4	2.0	2.6
LOW $r \approx .3$	0 0 0 c	RxRy	.048 (4,0)	.053 (4,0)	.061(16,0)	.068 (16,0)
	0 0 c c	RxRy	.052 (4,0)	.048 (4,0)	.047 (2,0)	.046 (0,0)
	0 0 0 c	RxR1y	.045 (4,0)	.045 (4,0)	.045 (4,0)	.045 (4,0)
	0 0 c c	RxR1y	.046 (4,0)	.045 (4,0)	.046 (4,0)	.045 (4,0)
	0 0 0 c	RxR2y	.048 (0,0)	.048 (0,0)	.048 (0,0)	.048 (0,0)
	0 0 c c	RxR2y	.048 (0,0)	.047 (0,0)	.048 (0,0)	.048 (0,0)
HIG H $r \approx .7$	0 0 0 c	RxRy	.050 (4,4)	.088(11,5)	.155(16,16)	.254 (16,16)
	0 0 c c	RxRy	.044 (4,4)	.045 (4,4)	.041 (4,3)	.034 (4,0)
	0 0 0 c	RxR1y	.031 (4,0)	.031 (4,4)	.031 (4,4)	.031 (4,4)
	0 0 c c	RxR1y	.031 (4,4)	.031 (4,4)	.031 (4,4)	.031 (4,4)
	0 0 0 c	RxR2y	.044 (4,0)	.044 (4,0)	.043 (4,0)	.043 (4,0)
	0 0 c c	RxR2y	.044 (4,0)	.044 (4,0)	.043 (4,0)	.043 (4,0)

Note. Numbers in parentheses are the number of times Type I error exceeded .055 and .075 in that condition where the maximum is 16. Rx indicates ranked X scores. Ry indicates ranked Y scores. R1y indicates ranked deviations of Y from the group Y mean. R2y indicates ranked deviations of Y from the group Y median.

The Type I error rate for RxRy increases as the magnitude of the treatment effect increases for the {0,0,0,c} pattern but not for the {0,0,c,c} pattern. The corresponding values for the methods based on deviations, RxR1y and RxR2y, have mean Type I error rates near .05. The simulations based on deviation scores with Type I error rates that surpassed the conservative robustness criteria are those based on SKLPT Y. The effects are more pronounced when there is a high correlation than when there is a low one.

The two new methods perform similarly in most of the simulations. The difference between them is predicted to be when there is a very skewed distribution. Table 4 presents the average Type I error rate of the {0,0,0,c} pattern for LEPTO Y and SKLPT Y distributions. The results are summed across shape of the X distribution and the additive constants. As expected the Type I error rate for the method based on deviations from the mean became problematic when the distribution is skewed. A symmetric leptokurtic distribution showed no elevation of Type I error with either of the new methods. The method based on the median is generally within acceptable bounds although it has more than a .06 error rate with the Skewed Leptokurtic, SKLPT Y.

Table 4. Comparing the two ranking alternatives across correlation, treatment effect and Y distribution.

	Correlation			
	LOW $r \approx .3$		HIGH $r \approx .7$	
	LEPTO Y	SKLPT Y	LEPTO Y	SKLPT Y
RxR1y	.044 (0,0)	.059 (32,0)	.023 (0,0)	.105 (36,36)
RxR2y	.044 (0,0)	.052 (0,0)	.021 (0,0)	.063 (36, 0)

Note. Numbers in parentheses are the number of times Type I error exceeded .055 and .075 respectively where the maximum is 36. Rx indicates Ranked X scores. R1y indicates Ranked Y deviations from the group Y mean. R2y indicates Ranked Y deviations from the group Y median.

Conclusion

Both of the problems associated with conducting a test of differences in slopes were replicated in the present study. Analysis of covariance on scores

that are not normally distributed have Type I error rates that systematically vary from the nominal value. If the distribution is leptokurtic the Type I error rate will be liberal and if it is platykurtic it will be conservative. It is difficult to determine if skew plays a role as most skewed distributions are also leptokurtic. The effect of shape is most clearly present when there is considerable shared variation in X and Y.

It is clearly the shape of the outcome measure rather than the covariate that results in manipulation of the Type I error rate. There is a small effect for the shape of X and little interaction between the shapes of X and Y. The complete set of 16 shapes is probably unrealistic in real world settings. The shapes of both X and Y are likely to be related to underlying characteristics of the sample chosen so that if Y is leptokurtic then X will likely also be somewhat leptokurtic. This would result in an accumulation of the major impact of the leptokurtic Y-scores and the minor impact of leptokurtic X scores to produce even more extreme Type I error elevation.

Tests of significance involving ranks rather than the original scores largely control the inflated Type I error rate although there appear to be unexplained differences in the error rate associated with ranking methods as a function of the underlying distribution. Specifically, the error rate is consistently higher for the Conover and Iman (1982) method when the SKLPT Y distribution was the source of the ranks. This trend for skewed distributions to produce larger Type I error rates even after being ranked was also found in Conover and Iman (1982) and Stephenson and Jacobson (1988).

The influence of additive treatment effects is shown to have the potentially serious inflation of Type I error noted by Headrick and Sawilowsky (2000). The effect was found where the additive effects tended to isolate one treatment group away from the remaining groups. Since the variance of ranks is based on the range of the ranks within the complete set, the separation of one group from a set of other groups will reduce the range and variance and produce a reduced slope. The effect appeared as the magnitude of the additive treatment effect increased. The beginning additive constant of .8 corresponds to a large effect in Cohen's (1988) terms. This effect showed no inflation of Type I error rate. Only as the additive

effect increased beyond this did the error rate become problematic.

Both of the proposed methods for testing slopes in the presence of potential additive effects reduced the Type I error rate to a generally acceptable level. The simulations that resulted in somewhat higher error rates were those with the most extreme distribution SKLPT Y. The method using deviations from the sample median was superior in controlling Type I error with SKLPT Y but was still somewhat elevated.

The two tests developed differ from others in that they are solely for testing the differences in slopes. There is no companion test for the presence of additive effects. A separate test such as that in Conover and Iman (1982) would need to be used for additive effects.

The ranking methods developed herein will have to be compared to other options to determine whether they have sufficient power to replace the traditional methods. The level of additive treatment effect used in the simulation is large and, at the upper end, may represent a level seen in relatively few experiments.

The experimenter should be able to anticipate this magnitude of effect. If the analysis of simple ranked scores as proposed by Conover and Iman (1982) is more powerful than the methods based on deviations the experimenter may choose to use simple ranks unless there is the expectation that very large additive treatment effects exist. However, if the power is equivalent the methods proposed herein should be preferred as they have more general Type I error control.

Lastly, the power of the tests using deviations from the mean and median need to be compared. While the median based method has superior Type I error control with the skewed leptokurtic distribution if it has less power the researcher may again want to determine if that condition within the outcome measure is likely to be present in the data and select accordingly.

References

- Atiquallah, M. (1964). The robustness of the covariance analysis of a one-way classification. *Biometrika*, 51, 365-372.
- Bradley, J. C. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

- Burnett, T. D., & Barr, D. R. (1977). A nonparametric analogy of analysis of covariance. *Educational and Psychological Measurement, 37*, 341-348.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conover, W. J., & Iman, R. L. (1982). Analysis of covariance using the rank transformation. *Biometrics, 38*, 715-724.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.
- Cronbach, L. J. & Snow, R. E. (1981). *Aptitudes and instructional methods: A handbook for research on interactions*. (2nd ed.). New York: Irvington.
- Dance, K. A., & Neufeld, R. W. J. (1988). Aptitude-Treatment Interaction research in the clinical setting: A review of attempts to dispel the "patient uniformity" myth. *Psychological Bulletin, 104*(2), 192-213.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521-532.
- Harwell, M. R., & Serlin, R. C. (1988). An experimental study of a proposed test of non-parametric analysis of covariance. *Psychological Bulletin, 104*, 268-281.
- Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika, 64*, 25-35.
- Headrick, T. C., & Sawilowsky, S. S. (2000). Properties of the rank transformation in factorial analysis of covariance. *Communications in Statistics- Simulation and Computation, 29*, 1059-1087.
- Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. NY: Wiley.
- Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86*(4), 721-735.
- Klockars, A. J., & Beretvas, S. N. (2001). Analysis of covariance and randomized block design with heterogeneous slopes. *The Journal of Experimental Education, 69*, 393-410.
- Klockars, A. J., & Moses, T. (2001). *Dealing with violations of the assumptions for aptitude X treatment interactions in ANCOVA*. Paper presented at the American Educational Research Association, Seattle.
- Klockars, A. J., Potter, N. S., & Beretvas, S. N. (1999). Power to detect additive treatment effects with randomized block and analysis of covariance designs. *The Journal of Experimental Education, 67*, 180-191.
- McClelland, G. H. & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376-390.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.
- Puri, M. L., & Sen, P. K. (1968). 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.
- Rheinheimer, D. C., & Penfield, D. A. (2001). The effects of type I error rate and power of the ANCOVA F test and selected alternatives under nonnormality and variance heterogeneity. *The Journal of Experimental Education, 69*, 373-391.
- Seaman, S., Algina, J., & Olejnik, S. F. (1985). Type I error probabilities and power of the rank and parametric ANCOVA procedures. *Journal of Educational Statistics, 10*, 345-367.
- Shirley, E. A. C. (1981). A distribution-free method for analysis of covariance based on ranked data. *Applied Statistics, 30*, 158-162.
- Stephenson, W. R., & Jacobson, D. (1988). A comparison of non-parametric analysis of covariance techniques. *Communication in Statistics: Simulation and Computation, 26*, 605-618.