Journal of Modern Applied Statistical Methods

Volume 4 | Issue 2 Article 11

11-1-2005

Testing For Aptitude-Treatment Interactions In Analysis Of Covariance And Randomized Block Designs Under Assumption Violations

Tim Moses
Educational Testing Service, Princeton, NJ, tmoses@ets.org

Alan Klockars
University of Washington, klockars@u.washington.edu

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 460-472

1538 – 9472/05/$95.00

Testing For Aptitude-Treatment Interactions In Analysis Of Covariance And Randomized Block Designs Under Assumption Violations

Tim Moses
Educational Testing Service
Princeton, NJ

Alan Klockars
University of Washington

This study compared the robustness of two analysis strategies designed to detect Aptitude-Treatment Interactions to two of their similarly-held assumptions, normality and residual variance homogeneity. The analysis strategies were the test of slope differences in analysis of covariance and the test of the Block-by-Treatment interaction in randomized block analysis of variance. With equal sample sizes in the treatment groups the results showed that residual variance heterogeneity has little effect on either strategy but nonnormality makes the test of slope differences liberal and the test of the Block-by-Treatment interaction conservative. With unequal sample sizes in the treatment groups the often-reported sample size-variance heterogeneity pairing is problematic for both strategies. The findings suggest that the randomized block strategy can be characterized as an overly-conservative alternative to the test of slope differences with respect to robustness.

Key words: Aptitude-treatment interactions, analysis of covariance, randomized block, nonnormality, variance heterogeneity, robustness

Introduction

One of the important issues in education is identifying when the effect of an instructional strategy depends on some individual difference variable (X) of the student. In their seminal work, Cronbach and Snow (1977) called these effects Aptitude-Treatment Interactions (ATIs). Two analysis approaches for identifying the presence of ATIs differ in terms of how they deal with an originally continuous X.

Tim Moses is an Associate Measurement Statistician at Educational Testing Service. He works primarily on the Advanced Placement Program. Tim completed his PhD in Educational Psychology at the University of Washington. Email him at tmoses@ets.org. Alan Klockars is Professor of Educational Psychology at the University of Washington. His research concerns multiple comparisons and, more recently, methods of conducting ATI research. Email him at klockars@u.washington.edu.

The first is a randomized block analysis of variance approach in which X is stratified into mutually exclusive subsets (Blocks). The second is a regression-based analysis of covariance approach in which the observed continuum of X is used. The question of interest is assessed with a test of the Block-by-Treatment interaction in the randomized block approach and a test of the homogeneity of regression coefficients in the analysis of covariance approach.

The randomized block and the analysis of covariance approaches have been compared in terms of relative power and apparent popularity. When their assumptions are met, both approaches control Type I error to an acceptable level, while the analysis of covariance strategy has superior power (Klockars & Beretvas, 2001; Cronbach & Snow, 1977; Pedhazur, 1997; Aiken & West, 1991). The power advantage is greatest when the randomized block strategy is based on a large number of blocks. In terms of popularity and familiarity for researchers, the randomized block strategy seems to have the advantage (Klockars & Beretvas, 2001; Keselman, Huberty, Lix, Olejnik, Cribbie, Donahue, Kowalchuk, Lowman, Petoskey, Keselman, & Levin, 1998; Maxwell, O'Callaghan, & Delaney, 1993). The purpose of the current study is to compare the two strategies in terms of a different criterion, their relative robustness to violations of assumptions about the normality and between-group variance homogeneity of the errors.

The two strategies make similar assumptions about the normality and variance homogeneity of the errors, but define error differently. In the randomized block design error is defined as the deviation of the scores from the mean of the Block-Treatment group. This mean reflects the outcome measure (Y) for all individuals in a treatment group who are categorized into the same block based on their X values. The error variance for the randomized block design is called the Subject/Block-by-Treatment Mean Square or S/BT. In analysis of covariance, error is defined as the difference between the Y scores and the predicted value based on the X value of the subject. The predicted value is from the best fitting least squares line for the treatment group. The error variance for analysis of covariance is called the adjusted subject Mean Square or the residual variance.

Research has considered the effects of nonnormality and variance heterogeneity on the robustness of the two strategies, but most of this work has been on the analysis of covariance strategy. None of this work has specifically compared the robustness of the two analysis strategies under the same assumption violations. This research suggests that the two assumption violations have different effects on the robustness of the analysis of covariance and randomized block strategies.

Nonnormality seems to have a stronger impact on the robustness of the analysis of covariance strategy than on the robustness of the randomized block strategy. The analysis of covariance strategy becomes liberal when the error distribution is heavy-tailed and conservative when it is light-tailed (Conover & Iman, 1982; Headrick & Sawilowsky, 2000; Klockars & Moses, 2002). The randomized block strategy is mildly affected by all but the most extreme conditions of nonnormality (Milligan, Wong & Thompson, 1987; Keselman, Carriere, & Lix, 1995).

The effect of variance heterogeneity on robustness depends on whether group sample sizes are equal. With equal sample sizes, variance heterogeneity has a negligible effect on the robustness of the analysis of covariance strategy (Dretzke, Levin & Serlin, 1982; Overton, 2001) and sometimes a negligible (Milligan, Wong & Thompson, 1987) or other times a liberal (Harwell, Rubinstein, Hayes & Olds, 1992) effect on the randomized block strategy. With unequal sample sizes, variance heterogeneity influences the robustness of the two strategies in the same way: when the group with the largest sample size has the smallest error variance (inverse pairing) both strategies are liberal, and when the group with the largest sample size has the largest error variance (direct pairing) both strategies are conservative. The current study considers the variance heterogeneity effect for equal and unequal sample sizes.

Finally, the effect of combined nonnormality and variance heterogeneity is interactive for the analysis of covariance strategy and additive for the randomized block strategy. For the analysis of covariance strategy, the two assumption violations slightly correct for each other (Deshon & Alexander, 1996). For the randomized block strategy, the two assumption violations are not interactive so that robustness depends mostly on the extent of variance heterogeneity (Keselman, et al., 1995; Harwell, et al., 1992).

It is difficult to recommend either analysis of covariance or randomized block as the more robust strategy when the errors are nonnormal and heterogeneous. Comparisons of the two strategies have focused on power when their assumptions are met and their popularity among researchers. The research that has evaluated the impact of the assumption violations on robustness has not directly compared the robustness of the two strategies. The current study was motivated by these concerns. The major questions are 1) for combinations of nonnormality and variance heterogeneity, which strategy is more robust? and 2) how will the relative robustness of these two strategies compare to what is known about their relative power?

Methodology

A Monte Carlo simulation study was conducted to investigate the relative robustness of the ATI analysis strategies. The null hypothesis of no ATI was true in all conditions. Empirical Type I error rates based on 10,000 iterations were generated for each condition. These empirical Type I error rates were then compared to the nominal Type I error rate of .05. Two treatment groups were used throughout the study. The following conditions were considered.

Analysis strategies

The standard analysis of covariance test of regression slope heterogeneity (Slopes) and the randomized block Block-by-Treatment Interaction analyses were compared. The randomized block strategy was evaluated using two (RB2) and four (RB4) blocks of X using median and quartile splits of the X variable based on the total sample. While the creation of the X blocks using of the total sample can create slightly unequal sample sizes even though the treatment group sizes are intended to be equal, the use of the total sample was preferred over the excessively liberal strategy of creating the X blocks within each separate treatment group (Myers & Well, 1995).

Assignment strategies

Two major strategies for assigning subjects to treatment conditions in randomized block and analysis of covariance are random assignment and assignment that utilizes subjects' X scores (Lomax, 2001; Myers & Well, 1995). When subjects are randomly assigned to treatments without regard for X, the randomized block strategy creates X blocks after treatments are administered (post hoc blocking). When subjects are assigned to treatments based on their X score, the randomized block strategy first creates the desired number of blocks in the total sample and then randomly assigns equal numbers of subjects to each of the treatments from each of the blocks. The approach of assigning subjects to treatments based on X and using the analysis of covariance is called systematic assignment (Dalton & Overall, 1977), meaning that subjects are first ranked on X and then assigned to treatments in a systematic pattern (i.e. 12211221…).

The consideration of analysis and assignment strategy resulted in six strategies to be investigated: analysis of covariance with random assignment, analysis of covariance with systematic assignment, RB2 and RB4 with random assignment (post hoc blocking) and RB2 and RB4 with assignment from the blocks.

Normality

Three shapes were used for X and the errors of Y, including a normal shape (skew=0, kurtosis=0), a skewed and heavy-tailed shape (skew=1, kurtosis=10) and an extremely skewed and heavy-tailed shape (skew=3, kurtosis=50). The shapes were generated with Fleishman's (1978) method (described below).

Variance Heterogeneity

Between-group variance heterogeneity was created to obtain a specified residual variance ratio of the treatment groups' residual variances based on the groups' deviations from their own regression lines. The variance heterogeneity considered in this study corresponds to how variance heterogeneity occurs in observed datasets (Oswald, Saad, & Sackett, 2000), meaning that groups differed more on their X-Y correlations and Y variances than on their X variances. The three considered residual variance ratios for the groups were 1/1, 3/1 and 15/1. For the conditions of unequal sample size, the residual variances were directly and inversely paired with the treatment group sample sizes.

To assess the correspondence of the considered levels of residual variance heterogeneity from treatment group regression lines to levels of variance heterogeneity from Block-by-Treatment Y means, Tables 1 and 2 give the ratios of the largest-to-smallest variances for the Block-by-Treatment cells of the RB2 and RB4 designs for all levels of assumption violations considered in this study. As analytical methods for deriving Y variances after forming categories on a correlated X variable are valid only for symmetric distributions (Maxwell & Delaney, 1993), the approach taken to produce the ratios in Tables 1 and 2 was simply to generate each distribution and residual variance heterogeneity combination in a total sample of 100,000 observations and then compute Y variances for the randomized block designs based on random assignment to treatment conditions (note that the variance ratios based on assignment from the X blocks are almost exactly equal).

Data were simulated so that the correlation was either .3 or .7 for one group. For the second group, the correlation was somewhat different from .3 or .7 so that, combined with a different Y variance, this second group's slope was equal the first group's slope while a desired level of variance heterogeneity was obtained.

Sample Size

Forty or eighty subjects per treatment group were used. The conditions of unequal sample size used forty subjects in one group and eighty in the other.

Data Generation Method

The following data generation method was used to create X and Y variables of desired distributions, variances and correlations while allowing for different assignment strategies to the treatment conditions.

1) N values of one standard normal variate, Z, were generated, where N was the total sample size based on two treatment groups that were intended to be of equal sample size.

2) X was created as a transformation of Z using Fleishman's (1978) method for generating nonnormal variables:

$$X = a + bZ + cZ^2 + dZ^3 \qquad (1)$$

The constants (a, b, c, and d) determined the first (mean), second (variance), third (skew) and fourth (kurtosis) moments of X. The values of the constants were derived to obtain the three distributions of interest in this study, where each distribution had a mean and variance of 0 and 1, respectively. The constants and resulting distributions are listed in Table 3.

3) An error variable for Y (E) was generated exactly as X was in steps 1 and 2. E had the same distribution as X.

4) Equal numbers of Xs and Es were randomly assigned to treatment groups 1 and 2. Depending on the particular strategy being studied, this involved either random assignment from the total available dataset (analysis of covariance and randomized block with post hoc blocking), random assignment from blocks of X (randomized block with assignment from the X blocks) or systematic assignment of the ranked X values to treatment groups (analysis of covariance with systematic assignment). The assignment strategies were the same in the unequal sample size conditions as in the equal sample size conditions, but after assignment one treatment group's sample size was reduced by ½, approximating an experimental study with massive loss of subjects from one of the two treatment groups.

5) Y was created as a function of X and E:

$$Y = \sigma_{Yk}[\rho_k X + (1- \rho_k{}^2)^{.5}E] \qquad (2),$$

where ρ_k was the desired X-Y correlation and σ_{Yk} is the desired standard deviation of Y for treatment group k. The values ρ_k and σ_{Yk} were determined for both treatment groups such that the two groups had the desired residual variance ratio and the null hypothesis of no slope differences was true. The values used are summarized in Table 4.

Table 1 Simulated ratios of largest-to-smallest Y variances in the Block-by-Treatment cells of the randomized block designs (XY correlation = .3, N=100,000).

Distribution of X and E		***Residual Variance Ratio***					
		1/1		3/1		15/1	
Skew	Kurtosis	RB2	RB4	RB2	RB4	RB2	RB4
0	0	1.0/1	1.1/1	2.9/1	3.1/1	14.5/1	15.4/1
1	10	1.0/1	1.1/1	3.0/1	3.2/1	14.9/1	16.3/1
3	50	1.1/1	1.3/1	3.0/1	3.4/1	14.3/1	15.3/1

Table 2 Simulated ratios of largest-to-smallest Y variances in the Block-by-Treatment cells of the randomized block designs (XY correlation = .7, N=100,000).

Distribution of X and E		***Residual Variance Ratio***					
		1/1		3/1		15/1	
Skew	Kurtosis	RB2	RB4	RB2	RB4	RB2	RB4
0	0	1.0/1	1.2/1	2.5/1	3.2/1	11.6/1	15.1/1
1	10	1.3/1	1.9/1	2.7/1	3.8/1	11.7/1	16.9/1
3	50	1.8/1	3.4/1	2.8/1	4.8/1	10.7/1	17.5/1

Table 3 Fleishman constants used to generate the variables

Skew	Kurtosis	a	b	c (=-a)	d
0	0	0	1	0	0
1	10	-.08772	.56426	.08772	.12621
3	50	-.17038	-.04789	.17038	.26005

Table 4 Correlations and standard deviations used to create levels of residual variance heterogeneity.

Residual Variance Ratio	ρ_k for Group 1	σ_{Yk} for Group 1	ρ_k for Group 2	σ_{Yk} for Group 2
		Low X-Y Relationship		
1/1	0.3	1	0.3	1
1/3	0.3	1	0.171871	1.679143
1/15	0.3	1	0.080933	3.706751
		High X-Y Relationship		
1/1	0.7	1	0.7	1
1/3	0.7	1	0.492773	1.421127
1/15	0.7	1	0.24535	2.853069

Programming

The programming for this study was done in SAS, using the CALL RANNOR (SAS Institute Inc., 1999a) routine for creating standard normal deviates and the PROC GLM (SAS Institute Inc., 1999b) function with Type III Sums of Squares for implementing the analysis strategies.

Assessing the Type I Error Rates

To identify the conditions with the strongest influence on Type I error, ANOVAs of the six manipulated variables and their two, three, four, five and six-way interactions were used. These ANOVAs were conducted separately for the equal and unequal sample size conditions. For equal sample sizes, the six independent variables (and their number of levels) were analysis strategy (3), assignment strategy (2), nonnormality (3), residual variance ratio (3), sample size (2) and overall X-Y correlation (2). For unequal sample sizes, the six independent variables (and their number of levels) were analysis strategy (3), assignment strategy (2), nonnormality (3), residual variance ratio (3), sample size-residual variance pairing (direct or inverse, 2) and overall X-Y correlation (2). Due to the stability of the empirical error rates, the two ANOVAs captured 100% of the variation in Type I error. Representative tables that illustrated the most important effects from the ANOVAs are also provided. The Type I error rates in these tables were considered as meaningfully different from the nominal .05 rate based on the criterion of +/- 2 standard errors range (.046-.054). Note that the +/- 2 standard error range is almost identical to Bradley's (1978) conservative range (.045-.055).

Results

Equal Sample Sizes

Table 5 presents the ten effects with the largest mean squares from the ANOVA of the error rates for equal sample sizes in the treatment groups. These ten effects accounted for 84.6% of the variation in Type I error rates. The two strongest effects were the analysis strategy and the analysis*normality interaction, accounting for 72.3% of the variation in Type I error. The assignment strategy's main effect and interactions with analysis, analysis*normality were also visible, but to a much smaller extent. Residual variance heterogeneity, XY correlation and sample size had small main effects.

Tables 6 and 7 illustrate the results of Type I error effects for equal treatment group sample sizes. These tables present the empirical Type I error rates for three analysis strategies across normality and residual variance heterogeneity ratios for the treatment group sample sizes of 40 and the overall XY correlation of .3. Table 6 shows the results for random assignment to treatment conditions. Table 7 shows the results when X was used to assign subjects to treatment conditions.

Table 5 The Ten Effects with the Largest Mean Squares, Equal Sample Sizes

Source	Sum of Squares (multiplied by 1,000)	df	Mean Square (multiplied by 1,000)
Analysis	5.644	2	2.822
Analysis*Normality	5.350	4	1.338
Analysis*Assignment	.456	2	.228
Analysis*N	.342	2	.171
Correlation	.148	1	.148
Assignment	.117	1	.117
N	.115	1	.115
ResVarHet	.204	2	.102
Analysis*Normality*Assignment	.335	4	.084
Correlation*Normality	.143	2	.072

Table 6 Type I Error Rates for Treatment Groups of 40, an XY correlation of .3, and Random Assignment to Treatment Conditions.

Distribution of X and E		***Residual Variance Ratio***								
		1/1			3/1			15/1		
Skew	Kurtosis	Slopes	RB2	RB4	Slopes	RB2	RB4	Slopes	RB2	RB4
0	0	.047	.048	.052	.046	.046	.051	.051	.051	.054
1	10	.054	.046	.051	.054	.045*	.051	.055*	.052	.056*
3	50	.068*	.044*	.044*	.058*	.042*	.042*	.066*	.036*	.038*

* Outside the +/- 2 standard error range (.046 to .054).

Table 7 Type I Error Rates for Treatment Groups of 40, an XY correlation of .3, and Assignment to Treatment Conditions Utilizing X.

Distribution of X and E		***Residual Variance Ratio***								
		1/1			3/1			15/1		
Skew	Kurtosis	Slopes	RB2	RB4	Slopes	RB2	RB4	Slopes	RB2	RB4
0	0	.050	.050	.051	.052	.050	.051	.053	.052	.056*
1	10	.056*	.046	.043*	.061*	.050	.045*	.071*	.053	.051
3	50	.069*	.041*	.034*	.076*	.040*	.034*	.088*	.039*	.033*

* Outside the +/- 2 standard error range (.046 to .054).

The most visible effect shown in Tables 6 and 7 is the effect of nonnormality on the analysis strategies. For the analysis of covariance strategy, increased nonnormality made Type I error liberal. For the randomized block strategies, increased nonnormality made Type I error conservative. The effect of nonnormality on the strategies was slightly larger when assignment to treatments used X (Table 7) than when assignment to treatments was random (Table 6). The effect of residual variance heterogeneity was very small when subjects are randomly assigned to treatments (Table 6), though RB4 was significantly liberal in two of the four sample size-correlation conditions where residual variance heterogeneity was most extreme. When subjects were assigned to treatments based on X, residual variance heterogeneity seemed to increase the liberalness of the analysis of covariance test when there was nonnormality. The results shown in Tables 6 and 7 were similar for the higher sample size and XY correlation.

Unequal Sample Sizes

Table 8 presents the ten effects with the largest mean squares from the ANOVA of the error rates for unequal sample sizes in the treatment groups. The mean squares were much larger when sample sizes were unequal, indicating that variations in Type I error are much greater for unequal sample sizes than for equal sample sizes. The ten effects in Table 8 accounted for 98.9% of the variation in Type I error rates. The two strongest effects were the residual variance-sample size pairing (direct or inverse) and this pairing in interaction with the levels of residual variance heterogeneity, 80.5% of the variation in Type I error. Many of the remaining ten effects in Table 8 also involved interactions with the residual variance-sample size pairing and the levels of residual variance heterogeneity. The main effects and interactions with analysis strategy accounted for less than 8% of total variability in Type I error, suggesting small but visible differences in the robustness of the three analysis strategies. The effects of assignment strategy, overall XY correlation, sample size and normality effects were very small when group sample sizes were unequal.

Tables 9 and 10 illustrate the effects of directly-paired sample sizes and residual variance ratios where the overall XY correlation was .3 and the assignment strategy was either random (Table 9) or based on X (Table 10). With equal residual variances (a residual variance ratio of 1/1), the slope test became liberal, RB2 became conservative and RB4 was not seriously affected. With residual variance heterogeneity, all Type I error rates became extremely conservative. The most conservative strategy was RB4. The RB2 and the analysis of covariance strategies had similar Type I error rates when distributions were normal. The combination of nonnormality and residual variance heterogeneity was visibly interactive for the analysis of covariance strategy, which became slightly less conservative as distributions became more nonnormal. In contrast, the effect of nonnormality was very small for RB2 and RB4. The error rates in Tables 9 and 10 are similar, suggesting that the assignment strategy used makes little difference when sample sizes are unequal.

Tables 11 and 12 illustrate the effects of inversely-paired sample sizes and residual variances. With no residual variance heterogeneity, nonnormality made the analysis of covariance test liberal, RB2 conservative, and had little effect on RB4. As residual variances became different all three analysis strategies became liberal, where the randomized block strategy based on four blocks (RB4) was the most liberal and the analysis of covariance and RB2 strategies had similarly-liberal Type I error rates. The combination of nonnormality and residual variance heterogeneity made all three strategies slightly less liberal than residual variance heterogeneity with normality. The error rates in Tables 11 and 12 are very similar, suggesting that assignment strategy makes little difference when sample sizes are unequal (like the results of direct pairing).

Table 8 The Ten Effects with the Largest Mean Squares, Unequal Sample Sizes

Source	Sum of Squares (multiplied by 1,000)	df	Mean Square (multiplied by 1,000)
Pairing	340.380	1	340.380
Pairing*ResVarHet	230.011	2	115.006
ResVarHet	55.485	2	27.743
Analysis*Pairing	23.954	2	11.977
Analysis	13.601	2	6.800
Analysis*Pairing*ResVarHet	18.513	4	4.628
Analysis*ResVarHet	11.645	4	2.911
Pairing*Normality	.447	2	2.236
Pairing*Correlation	.622	1	.622
Pairing*ResVarHet*Normality	2.362	4	.591

Table 9 Type I Error Rates for the Direct Pairing of Sample Size (80, 40) and Residual Variance, an XY correlation of .3, and Random Assignment to Treatment Conditions.

Distribution of X and E		***Residual Variance Ratio***								
		1/1			3/1			15/1		
Skew	Kurtosis	Slopes	RB2	RB4	Slopes	RB2	RB4	Slopes	RB2	RB4
0	0	.050	.050	.050	.021*	.021*	.012*	.008*	.008*	.003*
1	10	.050	.049	.051	.025*	.022*	.015*	.015*	.006*	.003*
3	50	.060*	.045*	.050	.040*	.020*	.016*	.026*	.006*	.002*

* Outside the +/- 2 standard error range (.046 to .054).

Table 10 Type I Error Rates for the Direct Pairing of Sample Size (80, 40) and Residual Variance, an XY correlation of .3, and Assignment to Treatment Conditions Utilizing X.

Distribution of X and E		***Residual Variance Ratio***								
		1/1			3/1			15/1		
Skew	Kurtosis	Slopes	RB2	RB4	Slopes	RB2	RB4	Slopes	RB2	RB4
0	0	.046	.049	.051	.023*	.019*	.012*	.009*	.008*	.004*
1	10	.050	.047	.051	.030*	.020*	.013*	.014*	.008*	.003*
3	50	.062*	.045*	.052	.042*	.022*	.017*	.032*	.006*	.002*

* Outside the +/- 2 standard error range (.046 to .054).

Table 11 Type I Error Rates for the Inverse Pairing of Sample Size (40, 80) and Residual Variance, an XY correlation of .3, and Random Assignment to Treatment Conditions.

Distribution of X and E		***Residual Variance Ratio***								
		1/1			3/1			15/1		
Skew	Kurtosis	Slopes	RB2	RB4	Slopes	RB2	RB4	Slopes	RB2	RB4
0	0	.049	.053	.050	.099*	.097*	.138*	.149*	.149*	.245*
1	10	.049	.045*	.052	.097*	.094*	.128*	.143*	.147*	.238*
3	50	.060*	.043*	.050	.092*	.085*	.114*	.114*	.138*	.210*

* Outside the +/- 2 standard error range (.046 to .054).

Table 12 Type I Error Rates for the Inverse Pairing of Sample Size (40, 80) and Residual Variance, an XY correlation of .3, and Assignment to Treatment Conditions Utilizing X.

Distribution of X and E		***Residual Variance Ratio***								
		1/1			3/1			15/1		
Skew	Kurtosis	Slopes	RB2	RB4	Slopes	RB2	RB4	Slopes	RB2	RB4
0	0	.049	.048	.052	.102*	.099*	.142*	.160*	.152*	.248*
1	10	.054	.047	.050	.097*	.100*	.127*	.147*	.153*	.240*
3	50	.061*	.048	.052	.092*	.081*	.111*	.131*	.145*	.215*

* Outside the +/- 2 standard error range (.046 to .054).

Conclusion

The purpose of the current study was to compare the robustness of two standard analysis strategies for detecting Aptitude-Treatment Interactions when two of their commonly-held assumptions were violated (nonnormal distributions and heterogeneous variances). The two strategies were the test for slope heterogeneity in analysis of covariance and the test of the Block-by-Treatment Interaction in randomized block analysis of variance. In addition, the strategies were evaluated based on two different assignment strategies, random assignment and assignment that utilized X.

The findings supported and extended the findings of previous studies that considered either the randomized block strategy (Milligan, Wong & Thompson, 1987; Keselman, Carrier & Lix, 1995; Harwell, Rubinstein, Hayes & Olds, 1992) or the analysis of covariance strategy (Conovar & Iman, 1982; Headrick & Sawilowsky, 2000; Klockars & Moses, 2002; Dretzke, Levin & Serlin, 1982; Overton, 2001; Deshon & Alexander, 1996; Conerly & Mansfield, 1988) separately. With equal sample sizes, the effect of nonnormality was much stronger than the effect of residual variance heterogeneity, causing the analysis of covariance strategy to get significantly liberal and the randomized block strategy to get significantly conservative. The effect of nonnormality was stronger when assignment to treatment groups was based on X than when assignment was random. With unequal sample sizes, the effect of residual variance heterogeneity was much stronger than the effect of nonnormality, causing the analysis strategies to get significantly conservative when residual variances were directly paired with sample sizes and liberal when residual variances were inversely paired with sample sizes. For unequal sample sizes the assignment strategy did not matter. Finally, for unequal sample sizes the combination of nonnormality and heterogeneous residual variances was interactive for the analysis of covariance strategy and slightly additive for the randomized block strategy. These findings suggest how the issue of robustness can contribute to several years of discussion on the relative merits of the randomized block and analysis of covariance strategies (Cox, 1957; Feldt, 1958; Cronbach & Snow, 1977; Aiken & West, 1991; Pedhazur, 1997; Lomax, 2001; Myers & Well, 1995; Klockars & Beretvas, 2001).

The magnitude of the effects of assumption violations on the robustness of the analysis strategies for equal sample sizes was somewhat smaller than expected. While heavy-tailed distributions did inflate the Type I error for the slope test, the inflation was rather small (up to about .09) given the extremely nonnormal distributions used. Two factors that kept Type I error from fluctuating too widely for extreme nonnormality were the assignment strategies, which made the treatment groups similar in the X distributions and therefore spread the extreme observations fairly evenly across the groups, and the use of a data generation method that created Y's nonnormality rather indirectly through adding nonnormality to X and E. Consistent with previous studies that used a similar data generation method (Conover & Iman, 1982; Luh & Gou, 2000), nonnormality has to be extreme and fairly unrealistic (Micceri, 1989) in order to see its effects on robustness with this data generation method.

The small effect of variance heterogeneity for the randomized block strategy with two blocks and equal sample sizes was surprising given the many studies that discuss the strong influence variance heterogeneity has on standard tests of means (Lix, Keselman, & Keselman, 1996) and interactions (Harwell, Rubinstein, Hayes, & Olds, 1992). However, many studies of the variance heterogeneity assumption focus much more on unequal sample sizes than on equal sample sizes (e.g. Milligan, Wong & Thompson, 1987; Keselman, Carriere & Lix, 1995), giving the impression that unequal sample sizes almost always accompany variance heterogeneity. For example, Milligan et al's study focuses almost completely on the effect of variance heterogeneity and unequal sample sizes, giving only a very quick mention of finding a negligible effect of heterogeneous variances when sample sizes were equal (p. 469). It is possible that the variance heterogeneity created from given levels of residual variance heterogeneity (Tables 1 and 2) was not large enough to impact the randomized block strategy with two blocks and equal sample sizes. In contrast to the randomized block strategy with two blocks, the randomized block strategy with four blocks resulted in greater levels of variance heterogeneity and did get liberal even when sample sizes were equal.

The explanations of the effects of the assumption violations on the analysis strategies are fairly well-known. Nonnormality makes treatment group slope estimates differ because of high-leverage observations that are extreme on both X and Y, resulting in inflated numerators of the F ratio. In addition, the standard errors of the slopes are smaller than they should be because the denominators of these standard errors use the sum of squares of X, which gets large as observations get more extreme. As the XY correlation increases, so does nonnormality's liberal effect on the test of slopes. For randomized block's tests of means, nonnormal Y's inflate standard deviations and standard errors, resulting in conservative tests. Nonnormal distributions can also affect mean estimates as well. In general, nonnormality has a stronger influence on sums of squares (standard deviations and standard errors) and sums of products (covariances) than it does on sums of raw data (means).

The effects of heterogeneous variances for equal and unequal sample sizes are also straightforward. The randomized block and analysis of covariance F tests use denominators that pool within-group variability across the groups. When sample sizes are equal, this pooling reasonably weights each group's variance equally. When sample sizes are unequal, the variance of the larger group gets weighted more heavily than that of the smaller group, which can over or underestimate random error and lead to conservative or liberal tests, respectively.

Given the effects of the assumption violations on the standard analysis strategies, many alternative strategies have been proposed. In fact, this study was motivated by a view of the randomized block strategy as an alternative strategy to the analyses of covariance strategy that might be more robust to nonnormal distributions. Other alternatives to the slope test include parametric alternative tests for heterogeneous residual variances (Deshon & Alexander, 1996; Overton, 2001; Dretzke, Levin & Serlin, 1982), ranking strategies for nonnormality (Conover & Iman, 1982; Headrick & Sawilowsky, 2000; Klockars & Moses, 2002), and combinations of strategies designed for addressing combinations of assumption violations (Luh & Guo, 2000, 2002). Given researchers' noted tendency to favor more familiar analysis strategies, the randomized block strategy was a practically-important method to evaluate. The findings of this study show that the randomized block strategy suffers from its own problems with respect to robustness. Given its relatively low power (Klockars & Beretvas, 2001) the randomized block strategy is probably best viewed as an overly conservative alternative to the slope strategy, along the same lines as ranked analysis of covariance. The low power of the randomized block test makes its recommendation difficult, especially given the complaints of low power in interaction studies (Aguinis & Pierce, 1998).

One interesting extension of this study would be to evaluate applications of alternative strategies that can address assumption violations within both the randomized block framework and the analysis of covariance framework. A combination of approaches like trimming/winsorizing observations or trimming test statistics for nonnormality and using a parametric alternative test that does not pool treatment group variances for variance heterogeneity has been shown to be effective for improving the robustness and power of tests of means (Keselman, Wilcox, Othman, Fradette, 2002; Luh & Guo, 1999; Keselman, Othman, Wilcox & Fradette, 2004). Some of these combinations of alternative strategies are applicable to tests of interactions. Along these same lines, some ways to trim observations and test statistics for nonnormality and also to use similar parametric alternative tests for heterogeneous residual variances have been considered for the analysis of covariance slope test (Luh & Guo, 2000, 2002). The relative effectiveness of these combinations of alternative strategies for analysis of covariance and randomized block strategies under the same degrees of assumption violations would be interesting to evaluate.

References

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods*, *1*, 296-314.

Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Bradley, J. C. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144-152.

Conerly, M. D. & Mansfield, E. R. (1988). An approximate test for comparing heteroscedastic regression models. *Journal of the American Statistical Association, 83*(403), 811-817.

Conover, W. J. & Iman, R. L. (1982). Analysis of covariance using the rank transformation. *Biometrics, 38,* 715-724.

Cox, D. R. (1957). The use of a concomitant variable in selecting an experimental design. *Biometrika, 44,* 150-158.

Cronbach, L. J. & Snow, R. E. (1977). *Aptitude and instructional methods.* Irvington, New York.

Dalton, S., & Overall, J. C. (1977). Nonrandom assignment in ANCOVA: The alternative ranks design. *The Journal of Experimental Education, 46,* 58-62.

DeShon, R. P.; Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods, 1*(3), 261-277.

Dretzke, B. J.; Levin, J. R.; Serlin, R. C. (1982). Testing for regression homogeneity under variance heterogeneity. *Psychological Bulletin, 91,* 376-383.

Feldt, L. S. (1958). A comparison of the precision of three experimental designs employing a concomitant variable. P*sychometrika, 23,* 335-353.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521-532.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing monte carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics, 17(4),* 315-339.

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.

Keselman, H. J., Carriere, K. C., & Lix, L. M. (1995). Robust and powerful nonorthogonal analyses. *Psychometrika, 60(3),* 395-418.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of Educational Research, 68*(3), 350-386.

Keselman, H. J., Othman, A. R., Wilcox, R. R. & Fradette, K. (2004). The new and improved two-sample t test. *Psychological Science, 15*(1), 47-51.

Keselman, H. J., Wilcox, R. R., Othman, A. R. & Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods, 1*(2), 288-309.

Klockars, A. J., & Beretvas, S. N. (2001). Analysis of covariance and randomized block design with heterogeneous slopes. *Journal of Experimental Education, 69*(4), 393-410.

Klockars, A. & Moses, T. (2002). Type I error rates for rank-based tests of homogeneity of slopes. *Journal of Modern Applied Statistical Methods, 1*(2), 452-460.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research, 66*(4), 579-619.

Lomax, R. G. (2001). *Statistical Concepts: A second course for education and the behavioral sciences (2nd edition).* Lawrence Erlbaum Associates: Mahwah, New Jersey.

Luh, W., & Guo, J. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normality and inequality of variances. *British Journal of Mathematical and Statistical Psychology*, *52*, 303-320.

Luh, W., & Guo, J. (2000). Approximate transformation trimmed mean methods to the test of simple linear regression slope equality. *Journal of Applied Statistics, 27*(7), 843-857.

Luh, W., & Guo, J. (2002). Using Johnson's transformation with approximate test statistics for the simple regression slope homogeneity. *The Journal of Experimental Education*, *71*(1), 69-81.

Maxwell, S. E., O'Callaghan, M. F., & Delaney, H. D. (1993). Analysis of covariance. In L. K. Edwards (Ed.), *Applied analysis of variance in behavioral science.* New York: Marcel Dekker.

Maxwell, S. E. & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*(1), 181-190.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156-166.

Milligan, G. W., Wong, D. S., & Thompson, P. A. (1987). Robustness properties of nonnorthogonal analysis of variance. *Psychological Bulletin, 101*(3), 464-470.

Myers, J. L. & Well, A. D. (1995). *Research design and statistical analysis.* Lawrence Erlbaum Associates: Hillsdale, New Jersey.

Oswald, F. L., Saad, S.; Sackett, P. R. (2000). The homogeneity assumption in differential prediction analysis: Does it really matter? *Journal of Applied Psychology, 85*(4), 536-541.

Overton, R. C. (2001). Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods, 6*(3), 218-233.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research.* Fort Worth, Texas.

SAS Institute Inc. (1999a). *SAS language reference: Dictionary, version 8.* Cary, NC: Author.

SAS Institute Inc. (1999b). *SAS/STAT* user's guide, version 8. Cary, NC: Author.