An Empirical Investigation of Four Tests for Interaction in the Context of Factorial Analysis of Covariance

Todd C. Headrick **George Vineyard**
Southern Illinois University

The Type I error and power properties of the parametric *F* test and three nonparametric competitors were compared in terms of a 3×4 factorial analysis of covariance layout. The focus of the study was on the test for interaction either in the presence and/or absence of main effects. A variety of conditional distributions, sample sizes, levels of variate and covariate correlation, and treatment effect sizes were investigated. The Puri and Sen (1969a) test had ultra-conservative Type I error rates and power losses when main effect(s) were present. The adjusted rank transform (Blair & Sawilowsky, 1990; Salter & Fawcett; 1993) had liberal Type I error rates when sampling was from moderate to extremely skewed distributions. The Hettmansperger (1984) chi-square test displayed acceptable Type I error rates for all distributions considered when sample sizes were ten or twenty. It is suggested that the Hettmansperger (1984) test be considered as an alternative to the parametric *F* test provided sample sizes are relatively equal and at least as large as ten.

The rank transform (RT) procedure was recommended as an alternative to the parametric procedure in multiple regression (Iman & Conover, 1979) and factorial analysis of covariance (Conover & Iman, 1981, 1982) when the assumption of population normality was violated. The steps for hypotheses testing using the RT consists of (a) replacing the raw scores with their respective rank order, (b) conducting the classical normal theory tests on the ranks, and (c) referring to the usual tables of percentage points.

Unfortunately, the parametric *F* test is not invariant with respect to monotone transformations (such as the RT). More specifically, the nonlinear nature of the RT may add (remove) interactions when such interactions were absent (present) in the original raw scores. For example, and contrary to the suggestions above, it has been demonstrated that the RT fails as a viable alternative to the parametric procedure with respect to tests for (a) interaction in factorial ANOVA (Blair, Sawilowsky, & Higgins, 1987; Thompson, 1991; 1993), (b) parallelism and interaction in factorial ANCOVA (Headrick, 1997; Headrick & Sawilowsky, 2000), and (c) additive and nonadditive models in multiple regression (Headrick & Rotou, 2000).

However, nonparametric tests can be substantially more powerful than the parametric *t* or *F* tests when the assumption of normality is violated. For example, the Mann-Whitney U-test has an impressive asymptotic relative efficiency of 3 relative to the two independent samples *t*-test when the population sampled from is exponential (Conover, 1999). Thus, nonparametric or distribution free tests should be considered when these tests demonstrate both (a) robustness with respect to Type I error and (b) a power advantage relative to the parametric test.

Sawilowsky (1990) reviewed ten competing tests for interaction in the context of factorial ANOVA and ANCOVA. On the basis of Type I error and power properties, three potential competitors to the parametric *F* test remain. These alternative nonparametric tests are: the adjusted RT procedure (Blair & Sawilowsky, 1990; Salter & Fawcett, 1993); the Hettmansperger (1984) procedure; and the Puri and Sen (1969a) procedure. It should be noted that the Hettmansperger (1984) and Puri and Sen (1969a) procedures consider only the total group regression slope. As such, it is assumed that the within group regression slopes are equal for these tests.

Purpose of the Study

The purpose of the study is to compare and contrast the relative Type I error and power properties of the parametric *F* test and the three aforementioned nonparametric procedures in the context of factorial ANCOVA using Monte Carlo techniques. From the results of the Monte Carlo study, a statement will be made with respect to the conditions under which any of the nonparametric tests are useful alternatives to the parametric *F* test. Because good nonparametric tests exist for main effects, the focus of this study is concerned with the test for interaction in the presence and/or absence of main effects.

Methodology

A completely randomized balanced design with fixed effects and one covariate was used. The structural model representing the design was:

$$Y_{ijk} = \mu + \beta(X_{ijk} - \overline{X}) + \alpha_i + \tau_j + (\alpha\tau)_{ij} + \varepsilon_{ijk}, \quad (1)$$

($i = 1,\ldots, I$; $j = 1,\ldots, J$; and $k = 1,\ldots, n$), where $I = 3$, $J = 4$, and $n = 5$, 10, and 20.
The levels of variate (Y_{ijk}) and covariate (X_{ijk}) correlation were ρ = 0, .3, .6, and .9. Note that the regression slope coefficient in (1), β, remained constant across groups.

The treatment effect patterns modeled in (1) were as follows:

1. The main effect τ nonnull, the main effect α null, and the interaction $(\alpha\tau)$ null:
   1(a). $\tau_1 = d$;
   1(b). $\tau_1 = \tau_2 = d$; and $\tau_3 = \tau_4 = -d$.
2. The main effects τ and α nonnull, and the interaction $(\alpha\tau)$ null:
   2(a). $\tau_2 = \alpha_1 = d$; and $\tau_3 = \alpha_2 = -d$; and
   2(b). $\tau_3 = \alpha_1 = d$; and $\tau_1 = \tau_2 = \tau_4 = \alpha_3 = -d$.
3. The $(\alpha\tau)$ interaction nonnull, and the main effects τ and α null:
   3(a). $(\alpha\tau)_{11} = (\alpha\tau)_{33} = d$; and $(\alpha\tau)_{13} = (\alpha\tau)_{31} = -d$;
   3(b). $(\alpha\tau)_{11} = (\alpha\tau)_{14} = (\alpha\tau)_{32} = (\alpha\tau)_{33} = d$; and
   $(\alpha\tau)_{12} = (\alpha\tau)_{13} = (\alpha\tau)_{31} = (\alpha\tau)_{34} = -d$.
4. The main effect τ and the $(\alpha\tau)$ interaction nonnull, and the main effect α null:
   4(a). $(\alpha\tau)_{11} = d$; and $(\alpha\tau)_{14} = -d$;
   4(b). $(\alpha\tau)_{11} = (\alpha\tau)_{12} = (\alpha\tau)_{31} = (\alpha\tau)_{32} = d$; and
   $(\alpha\tau)_{13} = (\alpha\tau)_{14} = (\alpha\tau)_{33} = (\alpha\tau)_{34} = -d$.
5. The main effects τ, α, and $(\alpha\tau)$ interaction are nonnull:
   5(a). $(\alpha\tau)_{21} = (\alpha\tau)_{24} = d$;
   5(b). $(\alpha\tau)_{11} = (\alpha\tau)_{12} = (\alpha\tau)_{32} = (\alpha\tau)_{33} = (\alpha\tau)_{34} = d$; and
   $(\alpha\tau)_{13} = (\alpha\tau)_{31} = (\alpha\tau)_{14} = -d$.

The treatment effect sizes (d) ranged from $d = 0.10\sigma$ to $d = 2.00\sigma$, where σ is the standard deviation of the population from which samples were drawn, in increments of 0.10σ. The null case was represented when $d = 0.00$ for all effects.

The parametric F statistic was calculated using the OLS sums of squares approach given in Winer, Brown, and Michels (1991) for factorial ANCOVA. The F statistic for interaction was then compared to the critical value from the usual F tables of percentage points.

The adjusted RT (adjRT) statistic was computed as follows: (a) the residuals were obtained from conducting a two-way ANOVA on the reduced model that included only the grouping variables; (b) the residuals and the covariate were then ranked without respect to group membership; and (c) the usual parametric ANCOVA procedure was conducted on the ranked residuals and ranked covariate to obtain the test statistic for interaction. This statistic was then compared to same critical F value as the parametric test.

The Hettmansperger (H) (1984) chi-square statistic was computed as follows: (a) the residuals (*RES*) were obtained from the regression of the variate on the reduced model that included the covariate and the grouping variables; (b) the residuals were then ranked (denoted as *RRES*) without respect to group membership; (c) the standardized ranked residuals (*SRRES*) were obtained according to the following equation: $SRRES = \sqrt{12}\left[\left(\frac{RRES}{N+1} - \frac{1}{2}\right)\right]$;

(d) the *SRRES* were then submitted to a two-way ANOVA; (e) the sums of squares for interaction term obtained from the ANOVA was then compared to the critical value from a chi-square distribution with $(I-1)(J-1)$ degrees of freedom (Hettmansperger, 1984).

The Puri and Sen (PS) (1969a) chi-square statistic was computed as follows: (a) the variate and covariate were ranked irrespective to group membership; (b) the cell means ($\overline{R}_{Yij}, \overline{R}_{Xij}$), column means ($\overline{R}_{Y.j}, \overline{R}_{X.j}$), row means ($\overline{R}_{Yi.}, \overline{R}_{Yi.}$), and overall grand means ($\overline{R}_{Y..}, \overline{R}_{X..}$) were then obtained from the ranks of the variate and covariate scores; (c) the *ij*-th difference score was then obtained as follows:

$$DIFF(R_{Yij}) = (\overline{R}_{Yij} - \overline{R}_{Y..}) - (\overline{R}_{Y.j} - \overline{R}_{Y..}) - (\overline{R}_{Yi.} - \overline{R}_{Y..}), \text{ and}$$

$$DIFF(R_{Xij}) = (\overline{R}_{Xij} - \overline{R}_{X..}) - (\overline{R}_{X.j} - \overline{R}_{X..}) - (\overline{R}_{Xi.} - \overline{R}_{X..});$$

(d) the *ij*-th residual scores were obtained from subtracting the predicted differences from the observed differences as follows: $RES_{ij} = DIFF(R_{Yij})) - \rho_{YX} DIFF(R_{Xij})$, where ρ_{YX} is the total group rank correlation coefficient between the variate and covariate; (e) the L_n statistic (Puri & Sen, 1969a) was then formulated as: $L_n = \mathbf{V}_{11}\sum_i \sum_j nRES_{ij}^2$, where $\mathbf{V}_{11}$ is the first element on the principal diagonal of the inverted variance-covariance matrix ($\mathbf{V}$); and (f) the computed value of L_n was subsequently compared to the critical value from a chi-square distribution with $(I-1)(J-1)$ degrees of freedom (Puri & Sen, 1969a).

Nine conditional distributions were simulated with zero means ($\mu = 0$), unit variances ($\sigma^2 = 1$), and varying degrees of γ_1, γ_2, γ_3, and γ_4. The distributions approximated in the simulation were: 1 = normal ($\gamma_1 = 0$, $\gamma_2 = 0$, $\gamma_3 = 0$, and $\gamma_4 = 0$), 2 = uniform ($\gamma_1 = 0$, $\gamma_2 = -6/5$, $\gamma_3 = 0$, and $\gamma_4 = 48/7$); 3 = Cauchy ($\gamma_1 = 0$, $\gamma_2 = 25$, $\gamma_3 = 0$, and $\gamma_4 = 4000$); 4 = double exponential ($\gamma_1 = 0$, $\gamma_2 = 3$, $\gamma_3 = 0$, and $\gamma_4 = 30$); 5 = logistic ($\gamma_1 = 0$, $\gamma_2 = 6/5$, $\gamma_3 = 0$, and $\gamma_4 = 48/7$); 6 = chi-square 8*df* ($\gamma_1 = 1$, $\gamma_2 = 3/2$, $\gamma_3 = 3$, and $\gamma_4 = 15/2$), 7 = chi-square 4*df* ($\gamma_1 = \sqrt{2}$, $\gamma_2 = 4$, $\gamma_3 = 6\sqrt{2}$, and $\gamma_4 = 30$), 8 = chi-square 2*df* ($\gamma_1 = 2$, $\gamma_2 = 6$, $\gamma_3 = 24$, and $\gamma_4 = 120$), and 9 = chi-square 1 *df* ($\gamma_1 = \sqrt{8}$, $\gamma_2 = 12$, $\gamma_3 = 48\sqrt{2}$, and $\gamma_4 = 480$). The preceding values of γ_1 (coefficient of skew), γ_2 (coefficient of kurtosis), γ_3, and γ_4 are the third, fourth, fifth, and sixth standardized cumulants from their associated probability density functions with the exception of the Cauchy distribution. Because the moments of a Cauchy *pdf* are infinite, the above values of γ_1, γ_2, γ_3, and γ_4 associated with this density were selected to yield a symmetric distribution with heavy tail-weight.

The steps employed for data generation follow the model developed by Headrick (2000). The Headrick (2000) procedure is an extension of the Headrick and Sawilowsky (1999, 2000) procedure for simulating multivariate nonnormal distributions. The Headrick (2000) procedure generated the Y_{ijk} and X_{ijk} for the *ij*-th group in (1) from the use of the following equations:

$$Y_{ijk} = c_0 + c_1 Y_{ijk}^{*} + c_2 Y_{ijk}^{*^2} + c_3 Y_{ijk}^{*^3} + c_4 Y_{ijk}^{*^4} + c_5 Y_{ijk}^{*^5} + \delta_{ij} d \text{, and} \qquad (2)$$

$$X_{ijk} = c_0 + c_1 X_{ijk}^{*} + c_2 X_{ijk}^{*^2} + c_3 X_{ijk}^{*^3} + c_4 X_{ijk}^{*^4} + c_5 X_{ijk}^{*^5}, \text{ where } Y_{ijk}^{*},\ X_{ijk}^{*} \sim \text{iid N}(0,1). \qquad (3)$$

The resulting Y_{ijk} and X_{ijk} were distributed with group means of $\delta_{ij}d$ and zero (respectively), unit variances, the desired values of γ_1, γ_2, γ_3, γ_4, and the desired within group correlation (ρ). In all experimental situations, Y_{ijk} and X_{ijk} followed the same distribution. The value of $\delta_{ij}d$ was the shift parameter added to the *ij*-th group for the treatment effect pattern considered. The coefficients c_0, c_1, c_2, c_3, c_4, and c_5 were determined by simultaneously solving equations 37, 38, 39, 40, 41, and 42 from Headrick (2000) for the desired values of γ_1, γ_2, γ_3, and γ_4. The values of Y^*_{ijk} and X^*_{ijk} in (2) and (3) were generated using the following algorithms:

$$Y^*_{ijk} = Z_{ijk}\rho^* + V_{ijk}\sqrt{1-\rho^{*2}}\text{, and} \tag{4}$$

$$X^*_{ijk} = Z_{ijk}\rho^* + W_{ijk}\sqrt{1-\rho^{*2}}\text{,} \tag{5}$$

where the Z_{ijk}, V_{ijk}, and W_{ijk} ~ iid N(0,1). The resulting Y^*_{ijk} and X^*_{ijk} were normally distributed with zero means, unit variances, and correlated at the intermediate value $\rho^{*2}_{Y^*_{ijk}X^*_{ijk}}$. The intermediate correlation, which is different from the desired post-correlation ($\rho_{Y_{ijk}X_{ijk}}$) except under conditional normality, was determined by solving equation 26 from Headrick (2000) for the bivariate case for $\rho^*_{Y^*X^*}$. When both variables follow the same distribution, equation 26 from Headrick (2000) can be expressed as follows:

$$\begin{aligned}\rho_{Y_{ijk}X_{ijk}} &= c_0^2 + 9c_4^2 + 2c_0(c_2+3c_4) + c_1^2\rho^{*2}_{Y^*_{ijk}X^*_{ijk}} + 6c_1c_3\rho^{*2}_{Y^*_{ijk}X^*_{ijk}} + 9c_3^2\rho^{*2}_{Y^*_{ijk}X^*_{ijk}} + 30\times \\ &c_1c_5\rho^{*2}_{Y^*_{ijk}X^*_{ijk}} + 90c_3c_5\rho^{*2}_{Y^*_{ijk}X^*_{ijk}} + 225c_5^2\rho^{*2}_{Y^*_{ijk}X^*_{ijk}} + 72c_4^2\rho^{*4}_{Y^*_{ijk}X^*_{ijk}} + 6c_3^2\rho^{*6}_{Y^*_{ijk}X^*_{ijk}} + 120c_3c_5\rho^{*6}_{Y^*_{ijk}X^*_{ijk}} \\ &+ 600c_3^2c_5^2\rho^{*6}_{Y^*_{ijk}X^*_{ijk}} + 24c_4^2\rho^{*8}_{Y^*_{ijk}X^*_{ijk}} + 120c_5^2\rho^{*10}_{Y^*_{ijk}X^*_{ijk}} + c_2^2(1+2\rho^{*4}_{Y^*_{ijk}X^*_{ijk}}) + 6c_2(c_4 + 4c_4\rho^{*4}_{Y^*_{ijk}X^*_{ijk}})\end{aligned} \tag{6}$$

Values of $c_0,\ldots,c_5$, and $\rho^*_{Y^*_{ijk}X^*_{ijk}}$ were solved for (6) using Mathematica (Version 4.0, 1999). The solution values of $c_0,\ldots,c_5$, the intermediate correlations ($\rho^{*2}_{Y^*_{ijk}X^*_{ijk}}$), and post-correlations ($\rho_{Y_{ijk}X_{ijk}}$) for the conditional distributions considered are compiled in Table 1.

The computer used to carry out the Monte Carlo was a Pentium III-based personal computer. All programming was done using Lahey Fortran 77 version 3.0 (1994), supplemented with various subroutines from RANGEN (Blair 1986). Using the chi-square and *F* tables of percentage points, the proportions of hypotheses rejected were recorded for the four different procedures. The nominal alpha level selected was .05. Twenty five thousand repetitions were simulated for each of the 9(type of distribution) × 4(level of correlation) × 21(effect size) × 10(treatment effect pattern) experiments.

Results

Adequacy of the Monte Carlo

For each repetition, separate values of ρ_{ij} and $\gamma_{1_{ij}}$, $\gamma_{2_{ij}}$, $\gamma_{3_{ij}}$, and $\gamma_{4_{ij}}$ for the variate and covariate for each of the *IJ* groups were computed. Average values of ρ_{ij} ($\bar{\rho}_{..}$), $\gamma_{1_{ij}}$ ($\bar{\gamma}_{1..}$), $\gamma_{2_{ij}}$ ($\bar{\gamma}_{2..}$), $\gamma_{3_{ij}}$ ($\bar{\gamma}_{3..}$), and $\gamma_{4_{ij}}$ ($\bar{\gamma}_{4..}$) were obtained by averaging the ρ_{ij}, $\gamma_{1_{ij}}$, $\gamma_{2_{ij}}$, $\gamma_{3_{ij}}$, and $\gamma_{4_{ij}}$ across the *IJ* groups. The values of $\bar{\rho}_{..}$, $\bar{\gamma}_{1..}$, $\bar{\gamma}_{2..}$, $\bar{\gamma}_{3..}$, and $\bar{\gamma}_{4..}$ were subsequently averaged across 25,000 (replications) × 21 (effect size) situations in the first treatment effect pattern for each conditional distribution. The average values of $\bar{\gamma}_{1..}$, $\bar{\gamma}_{2..}$, $\bar{\gamma}_{3..}$, and $\bar{\gamma}_{4..}$ were then further averaged across the four levels of correlation. The overall

Table 1. Values of constants ($c_0,\ldots,c_5$) used in equation (3), population correlations ($\rho_{Y_{ijk}X_{ijk}}$), and intermediate correlations ($\rho^{*2}_{Y^*X^*}$) to simulate and correlate the desired conditional distributions (Dist).

Dist	c_0	c_1	c_2	c_3	c_4	c_5	$\rho_{Y_{ijk}X_{ijk}}$	$\rho^{*2}_{Y^*X^*}$
1	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	.00	.000000
							.30	.300000
							.60	.700000
							.90	.900000
2	0.000000	1.347438	0.000000	-0.140177	0.000000	0.001808	.00	.000000
							.30	.326197
							.60	.634118
							.90	.913613
3	0.000000	0.306093	0.000000	0.184686	0.000000	0.001132	.00	.000000
							.30	.374236
							.60	.683980
							.90	.929263
4	0.000000	0.727709	0.000000	0.096303	0.000000	-0.002232	.00	.000000
							.30	.309371
							.60	.612882
							.90	.905531
5	0.000000	0.879467	0.000000	0.040845	0.000000	-0.000405	.00	.000000
							.30	.302233
							.60	.603260
							.90	.901368
6	-0.163968	0.950794	0.165391	0.007345	-0.000474	0.000014	.00	.000000
							.30	.311431
							.60	.612677
							.90	.904625
7	-0.227508	0.900716	0.231610	0.015466	-0.001367	0.000055	.00	.000000
							.30	.322263
							.60	.624030
							.90	.908552
8	-0.307740	0.800560	0.318764	0.033500	-0.003675	0.000159	.00	.000000
							.30	.341958
							.60	.643339
							.90	.914879
9	-0.397725	0.621071	0.416907	0.068431	-0.006394	0.000044	.00	.000000
							.30	.376853
							.60	.673908
							.90	.924127

averages of $\overline{\gamma}_1$, $\overline{\gamma}_2$, $\overline{\gamma}_3$, $\overline{\gamma}_4$, and $\overline{\rho}$ are listed in Table 2 and Table 3, respectively. Inspection of Tables 2 and 3 indicate that the Headrick (2000) procedure produced excellent agreement between $\overline{\gamma}_1$, $\overline{\gamma}_2$, $\overline{\gamma}_3$, $\overline{\gamma}_4$, and $\overline{\rho}$ and the population parameters considered.

The Type I error and power analyses are compiled in Tables 4 through 13. The column entries from left to right denote (a) the test statistic, (b) the standardized treatment effect size "d", and (c) the proportion of rejections for the four different tests of interaction under the various levels of variate and covariate correlation and the other parameters considered.

Type I Error

Normal Distribution: The Type I error rates for the competing procedures are compiled in Tables 4, 6, and 8, for $n=5$, 10, 20, and treatment pattern 2(b). This particular effect pattern is reported because the commonly used rank transform test statistic (Conover & Iman, 1981) under these circumstances is not

Table 2. Average values of $\gamma_1(\bar{\gamma}_1)$, $\gamma_2(\bar{\gamma}_2)$, $\gamma_3(\bar{\gamma}_3)$, and $\gamma_4(\bar{\gamma}_4)$ simulated by the Headrick (2000) procedure. The average values ($\bar{\gamma}_1$, $\bar{\gamma}_2$, $\bar{\gamma}_3$, $\bar{\gamma}_4$) listed below were based on a sample size is $n=20$.

Distribution	Population parameter (γ_1, γ_2, γ_3, γ_4)			
1	$\gamma_1=0$	$\gamma_2=0$	$\gamma_3=0$	$\gamma_4=0$
Variate (Y)	$\bar{\gamma}_1=0.000124$	$\bar{\gamma}_2=-0.000284$	$\bar{\gamma}_3=0.001073$	$\bar{\gamma}_4=-0.001339$
Covariate (X)	$\bar{\gamma}_1=-0.000084$	$\bar{\gamma}_2=0.000452$	$\bar{\gamma}_3=0.000795$	$\bar{\gamma}_4=0.002845$
2	$\gamma_1=0$	$\gamma_2=-6/5$	$\gamma_3=0$	$\gamma_4=48/7$
Variate (Y)	$\bar{\gamma}_1=0.000005$	$\bar{\gamma}_2=-1.200004$	$\bar{\gamma}_3=0.0000238$	$\bar{\gamma}_4=6.857894$
Covariate (X)	$\bar{\gamma}_1=0.000039$	$\bar{\gamma}_2=-1.200163$	$\bar{\gamma}_3=0.0001685$	$\bar{\gamma}_4=6.853492$
3	$\gamma_1=0$	$\gamma_2=25$	$\gamma_3=0$	$\gamma_4=4000$
Variate (Y)	$\bar{\gamma}_1=-0.001318$	$\bar{\gamma}_2=24.975520$	$\bar{\gamma}_3=-.3386690$	$\bar{\gamma}_4=3958.22114$
Covariate (X)	$\bar{\gamma}_1=0.000290$	$\bar{\gamma}_2=24.941770$	$\bar{\gamma}_3=-0.799517$	$\bar{\gamma}_4=3988.30400$
4	$\gamma_1=0$	$\gamma_2=3$	$\gamma_3=0$	$\gamma_4=30$
Variate (Y)	$\bar{\gamma}_1=0.000342$	$\bar{\gamma}_2=2.999848$	$\bar{\gamma}_3=0.014447$	$\bar{\gamma}_4=30.010830$
Covariate (X)	$\bar{\gamma}_1=0.000032$	$\bar{\gamma}_2=3.000327$	$\bar{\gamma}_3=0.004328$	$\bar{\gamma}_4=30.006732$
5	$\gamma_1=0$	$\gamma_2=6/5$	$\gamma_3=0$	$\gamma_4=48/7$
Variate (Y)	$\bar{\gamma}_1=0.000224$	$\bar{\gamma}_2=1.199900$	$\bar{\gamma}_3=.004258$	$\bar{\gamma}_4=6.846827$
Covariate (X)	$\bar{\gamma}_1=0.000034$	$\bar{\gamma}_2=1.200087$	$\bar{\gamma}_3=.001478$	$\bar{\gamma}_4=6.858595$
6	$\gamma_1=1$	$\gamma_2=3/2$	$\gamma_3=3$	$\gamma_4=15/2$
Variate (Y)	$\bar{\gamma}_1=1.000071$	$\bar{\gamma}_2=1.500197$	$\bar{\gamma}_3=3.001597$	$\bar{\gamma}_4=7.496629$
Covariate (X)	$\bar{\gamma}_1=0.999992$	$\bar{\gamma}_2=1.500053$	$\bar{\gamma}_3=3.005218$	$\bar{\gamma}_4=7.538564$
7	$\gamma_1=\sqrt{2}$	$\gamma_2=3$	$\gamma_3=6\sqrt{2}$	$\gamma_4=30$
Variate (Y)	$\bar{\gamma}_1=1.414330$	$\bar{\gamma}_2=3.000764$	$\bar{\gamma}_3=8.489000$	$\bar{\gamma}_4=29.978800$
Covariate (X)	$\bar{\gamma}_1=1.413904$	$\bar{\gamma}_2=3.001067$	$\bar{\gamma}_3=8.484897$	$\bar{\gamma}_4=30.004765$
8	$\gamma_1=2$	$\gamma_2=6$	$\gamma_3=24$	$\gamma_4=120$
Variate (Y)	$\bar{\gamma}_1=2.000254$	$\bar{\gamma}_2=6.002129$	$\bar{\gamma}_3=24.008980$	$\bar{\gamma}_4=119.868700$
Covariate (X)	$\bar{\gamma}_1=1.999989$	$\bar{\gamma}_2=6.000573$	$\bar{\gamma}_3=24.010045$	$\bar{\gamma}_4=120.158647$
9	$\gamma_1=\sqrt{8}$	$\gamma_2=12$	$\gamma_3=48\sqrt{2}$	$\gamma_4=480$
Variate (Y)	$\bar{\gamma}_1=2.828878$	$\bar{\gamma}_2=12.003800$	$\bar{\gamma}_3=67.884840$	$\bar{\gamma}_4=479.035600$
Covariate (X)	$\bar{\gamma}_1=2.827901$	$\bar{\gamma}_2=12.000050$	$\bar{\gamma}_3=67.885672$	$\bar{\gamma}_4=480.001874$

asymptotically chi-squared (Thompson, 1991, 1993) and is liberal for even small samples (Headrick, 1997; Headrick & Sawilowsky, 2000).

As expected, the parametric F test maintained Type I error rates close to nominal alpha and were within the closed interval of $\alpha \pm 1.96\sqrt{\alpha(1-\alpha)/25000}$. This occurred across all treatment conditions, sample sizes, and levels of variate/covariate correlation.

The adjRT also generated acceptable Type I error rates. Inspection of Tables 4, 6, and 8 indicates that the Type I error rates were similar to the parametric F test. With respect to the H test, inspection of Tables 6 and 8 indicates that this test maintained appropriate Type I error rates for sample sizes of $n=10$ and $n=20$. However, for $n=5$, inspection of Table 4 indicates that the H test generated liberal Type I error rates. For example, with an effect size of $d=0.80$, the Type I error rates were approximately .060 across all levels of variate/covariate correlation.

Table 3. Average values of variate and covariate correlation ($\bar{\rho}$) simulated by the Headrick (2000) procedure. The value ρ denotes the population correlation. The average values ($\bar{\rho}$) listed below were based on a sample size is $n=20$.

		Distribution								
n	ρ	1	2	3	4	5	6	7	8	9
20	.00	.000	.000	-.001	.000	.000	.000	.001	.000	-.000
	.30	.300	.299	.300	.301	.300	.299	.300	.300	.301
	.60	.600	.601	.602	.599	.600	.598	.600	.599	.600
	.90	.900	.899	.901	.900	.900	.901	.899	.900	.900

The PS test became conservative when either one or both main effects were present. *Ceteris paribus*, the stronger the nonnull main effect(s) the more conservative the Type I error rates became. These conservative Type I error rates occurred across all levels of variate and covariate correlation. For example, with an effect size of $d=0.80$, inspection of Table 4 indicates that the Type I error rates were .001, .000, and .000 across the three levels of variate/covariate correlation. The PS procedure maintained Type I error rates close to nominal alpha only when *both* main effects were null.

Nonnormal Distributions: Type I error rates are compiled in Tables 10 and 12 for some of the nonnormal distributions considered. The approximate distributions reported in these tables are the chi-square 1*df* and Cauchy. These distributions are reported because previous empirical investigations demonstrated that Type I error inflations associated with the rank transform test statistic (Conover & Iman, 1981) were most severe under extreme departures from normality (Headrick, 1997; Headrick & Sawilowsky, 2000).

The parametric F test was slightly conservative under the nonnormal conditional distributions reported. For example, with an effect size of $d=1.30$, variate/covariate correlation of $r=.30$, an inspection of Table 12 indicates that the Type I error rate was .040 when the conditional distribution was approximate Cauchy.

The adjRT generated inflated Type I error rates when the conditional distribution considered was skewed (e.g., chi-square 1*df* or 2*df*). For example, with an effect size of $d=0.80$, a variate/covariate correlation of $r=.90$, inspection of Table 10 indicates that the Type 1 error rate for the adjRT was .076. In general, increases in skew i.e., chi-square 4*df*, chi-square 2*df*, chi-square 1*df* were associated with increases in Type I error inflation for the adjRT.

The H test maintained appropriate Type I error rates for all nonnormal conditional distributions considered when sample sizes were $n=10$ and $n=20$. When samples were $n=5$, the H test generated liberal Type I error rates. The inflated Type I error rates were similar to those error rates generated under conditional normality.

As with the standard normal case, the PS test generated ultra-conservative Type I error rates when main effects were present. For example, with an effect size $d=0.80$ and a variate/covariate correlation of $r=.60$, inspection of Table 12 indicates that the Type I error rate was .000. This occurred for *all* nonnormal distributions considered in this study.

Power Analysis

Normal Distribution: Power analyses for the competing procedures are compiled in Tables 5, 7, and 9, for $n=5$, 10, 20, and treatment pattern 5(a). This effect pattern is reported because under these conditions the usual rank transform statistic has been demonstrated to display severe power losses (Headrick, 1997; Headrick & Sawilowsky, 2000).

As expected, the F test displayed a power advantage over the three nonparametric competitors when the conditional distribution was standard normal. Specifically, the F test was substantially more powerful than the PS test when both main effects became increasingly nonnull. Although the F test was more powerful than the H test when sample sizes were $n=10$ and $n=20$, the H test held a slight power advantage over the adjRT. When sample sizes were $n=5$, inspection of Table 5 indicates that the H test

Table 4. Type I error results for the test of interaction. The sampling distribution was standard normal. The sample size was $n=5$. Both main effects were nonnull. The Type I error rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (d)	0.3	0.6	0.9
F	0.30	.051	.052	.050
adjRT		.052	.053	.051
H		.057	.060	.060
PS		.023	.020	.004
F	0.80	.050	.051	.050
adjRT		.052	.053	.049
H		.059	.056	.058
PS		.001	.000	.000
F	1.30	.052	.047	.052
adjRT		.052	.050	.051
H		.059	.060	.061
PS		.000	.000	.000

Table 5. Power analysis for the test of interaction when sampling was from a standard normal distribution. The sample size was $n=5$. Both main effects were nonnull. The rejection rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (d)	0.3	0.6	0.9
F	0.30	.062	.067	.110
adjRT		.062	.066	.100
H		.069	.077	.121
PS		.055	.056	.081
F	0.80	.145	.202	.622
adjRT		.143	.187	.531
H		.159	.217	.632
PS		.106	.132	.315
F	1.30	.359	.507	.983
adjRT		.349	.473	.954
H		.372	.517	.983
PS		.211	.272	.272

was rejecting at a higher rate than F test. For example, with an effect size of $d=0.80$, a variate/covariate correlation of $r=.30$, inspection of Table 5 indicates that the H test had a rejection rate of .16 while the F test was rejecting at a rate of .145. This higher rejection rate is attributed to the liberal nature of the Type I error rates that were associated with the H test when $n=5$.

Nonnormal Distributions: In general, when departures from normality were small (e.g., approximate logistic) to moderate (e.g., approximate chi-square 8*df*) the F test rejected at rates slightly less than the Hettmansperger and adjRT procedures. The power advantages in favor of either the H or adjRT tests were contingent on the conditional distribution considered and the other parameters being simulated. It should be noted that the power advantages in favor either the H test or adjRT test were marginal. On the other hand, when the conditional distribution was approximate uniform the parametric F test held a slight advantage over the nonparametric procedures.

When the conditional distributions were extremely skewed and/or heavy tailed, both the adjRT and H tests held large power advantages over the F test. Further, when the adjRT test generated reasonable Type I error rates, the adjRT displayed some power advantages over the other competing nonparametric procedures. For example, inspection of Table 13 indicates that when the conditional distribution was approximate Cauchy, an effect size of $d=0.80$, and a variate/covariate correlation of $r=.60$, the adjRT

Table 6. Type I error results for the test of interaction. The sampling distribution was standard normal. The sample size was $n = 10$. Both main effects were nonnull. The Type I error rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (*d*)	0.3	0.6	0.9
F	0.30	.053	.052	.049
adjRT		.054	.051	.049
H		.052	.055	.053
PS		.027	.019	.003
F	0.80	.050	.050	.050
adjRT		.051	.049	.051
H		.053	.052	.053
PS		.006	.001	.000
F	1.30	.050	.048	.050
adjRT		.051	.048	.051
H		.054	.051	.054
PS		.000	.000	.000

Table 7. Power analysis for the test of interaction when sampling was from a standard normal distribution. The sample size was $n = 10$. Both main effects were nonnull. The rejection rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (*d*)	0.3	0.6	0.9
F	0.30	.078	.088	.208
adjRT		.077	.087	.179
H		.076	.089	.201
PS		.069	.077	.155
F	0.80	.294	.418	.951
adjRT		.284	.386	.911
H		.288	.402	.943
PS		.232	.306	.779
F	1.30	.715	.879	1.000
adjRT		.693	.848	.999
H		.697	.863	1.000
PS		.531	.683	.987

was rejecting at a rate of .942 whereas the H test was rejecting at a rate of .844. Power comparisons between these two tests were not considered where the adjRT generated liberal Type error rates (e.g., approximate chi-square 2*df* or chi-square 1*df*). When the conditional distributions were approximately chi-square 2*df* or chi-square 1*df*, the H test was a much more powerful than the parametric *F*. For example, when sampling was from an approximate chi-square distribution 1*df*, $d = 0.80$, a variate/covariate correlation of $r = .30$, inspection of Table 11 indicates that the H test was rejecting at a rate of .731 while the F test was rejecting at a rate of only .326.

The PS procedure held a power advantage over the H and adjRT tests only when *both* main effects were either weak or null. Otherwise, the PS test statistic had the problem of power loss when juxtaposed to either the H or the adjRT tests as the magnitude of the main effect(s) increased. For example, when sampling was from an approximate chi-square distribution 1*df*, $d = 0.30$, a variate/covariate correlation of $r = .30$, inspection of Table 11 indicates that the PS test was rejecting at a rate of .182 while the H test was rejecting at a rate of .148. However, when the effect size increased from $d = .30$ to $d = 0.80$, the H test was rejecting at a rate of .731 while the PS was rejecting at a rate of only .524. This pattern of power loss associated with the PS test was consistent across all nonnormal distributions considered in this study.

Table 8. Type I error results for the test of interaction. The sampling distribution was standard normal. The sample size was $n = 20$. Both main effects were nonnull. The Type I error rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (d)	0.3	0.6	0.9
F	0.30	.050	.050	.049
adjRT		.050	.050	.050
H		.051	.052	.050
PS		.028	.019	.003
F	0.80	.051	.049	.052
adjRT		.052	.052	.051
H		.052	.052	.052
PS		.001	.000	.000
F	1.30	.050	.050	.050
adjRT		.050	.049	.048
H		.052	.051	.052
PS		.000	.000	.000

Table 9. Power analysis for the test of interaction when sampling was from a standard normal distribution. The sample size was $n = 20$. Both main effects were nonnull. The rejection rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (d)	0.3	0.6	0.9
F	0.30	.109	.133	.410
adjRT		.105	.127	.360
H		.105	.131	.393
PS		.099	.119	.328
F	0.80	.596	.775	1.000
adjRT		.569	.734	1.000
H		.573	.754	1.000
PS		.505	.661	.994
F	1.30	.976	.998	1.000
adjRT		.968	.996	1.000
H		.969	.997	1.000
PS		.920	.978	1.000

Discussion

The PS test is computationally arduous. Further, the results of this study indicate that this test had the problems of ultra-conservative Type I error rates and power loss when main effects were nonnull. Toothaker and Newman (1994) found similar results with respect to the PS test in the context of factorial ANOVA. Thus, it is recommended that this procedure not be considered as a viable alternative to the parametric F test in factorial ANCOVA.

It is possible to base the PS statistic on normal or expected normal scores instead of the ranks (Puri & Sen, 1969a). And, this *might* correct the problem of ultra-conservative Type I error rates. However, additional nonlinear transformations present the problem with respect to the correct interpretation of the statistical results in terms of the original metric.

The adjRT is arguably the simplest of the three nonparametric procedures to compute. However, because the adjRT has the problem of liberal Type I error rates when the distributions possess moderate to extreme skewness, it is also recommended that the adjRT procedure not be used in place of the parametric F test.

Table 10. Type I error results for the test of interaction. The sampling distribution was an approximate chi-square distribution with 1degree of freedom. The sample size was $n = 10$. Both main effects were nonnull. The Type I error rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (*d*)	0.3	0.6	0.9
F	0.30	.046	.043	.045
adjRT		.069	.068	.075
H		.053	.048	.048
PS		.010	.006	.001
F	0.80	.044	.047	.047
adjRT		.067	.072	.076
H		.051	.050	.049
PS		.004	.001	.000
F	1.30	.045	.046	.046
adjRT		.070	.070	.073
H		.052	.049	.049
PS		.000	.000	.000

Table 11. Power analysis for the test of interaction when sampling was from was an approximate chi-square distribution with 1degree of freedom. The sample size was $n = 10$. Both main effects were nonnull. The rejection rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (*d*)	0.3	0.6	0.9
F	0.30	.075	.086	.221
adjRT		.197	.262	.679
H		.148	.182	.522
PS		.182	.243	.597
F	0.80	.326	.462	.947
adjRT		.815	.926	.999
H		.731	.838	.999
PS		.524	.672	.973
F	1.30	.739	.881	.999
adjRT		.991	.999	1.000
H		.981	.995	1.000
PS		.762	.885	.998

The H chi-square test maintained appropriate Type I error rates for *all* conditional distributions considered in this study when sample sizes were at least as large as $n = 10$. Thus, the H test could be considered as an alternative to the parametric *F* test for interaction provided the within group sample sizes are relatively equal and *at least as large* as $n = 10$. This recommendation is made in view of the large power advantage that the H test had over the *F* test when the conditional distributions were contaminated with outliers and/or possessed extreme skewness.

Table 12. Type I error results for the test of interaction. The sampling distribution was an approximate Cauchy distribution. The sample size was $n = 10$. Both main effects were nonnull. The Type I error rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (d)	0.3	0.6	0.9
F	0.30	.043	.045	.044
adjRT		.054	.053	.058
H		.046	.050	.050
PS		.008	.005	.001
F	0.80	.044	.044	.046
adjRT		.053	.055	.056
H		.045	.048	.048
PS		.000	.000	.000
F	1.30	.040	.044	.045
adjRT		.053	.052	.056
H		.045	.047	.048
PS		.000	.000	.000

Table 13. Power analysis for the test of interaction when sampling was from was an approximate Cauchy distribution. The sample size was $n = 10$. Both main effects were nonnull. The rejection rates were based on 25,000 repetitions and a nominal alpha level of $\alpha = .05$.

		Level of Correlation		
Test	Effect Size (d)	0.3	0.6	0.9
F	0.30	.075	.096	.244
adjRT		.163	.235	.699
H		.130	.173	.495
PS		.155	.220	.632
F	0.80	.346	.489	.946
adjRT		.801	.942	1.00
H		.712	.844	.999
PS		.563	.754	.993
F	1.30	.750	.884	.999
adjRT		.993	.999	1.00
H		.981	.996	1.00
PS		.804	.934	.999

References

Blair, R. C. (1987). *RANGEN*. Boca Raton, FL: IBM.

Blair, R. C., & Sawilowsky, S. S. (1990, April). *A test for interaction based on the rank transform*. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Blair, R. C., Sawilowsky, S. S., & Higgins, J. J. (1987). Limitations of the rank transform in tests of interactions. *Communications in Statistics*: *Simulation and Computation*, *16*, 1133-1145.

Conover, W.J. (1999). *Practical nonparametric statistics* (3rd). New York: Wiley.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*, 124-133.

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

Headrick, T. C. (1997). Type I error and power of the rank transform analysis of covariance (ANCOVA) in a 3×4 factorial layout. Unpublished doctoral dissertation, Wayne State University, Detroit, MI.

Headrick, T. C., (2000). Simulating univariate and multivariate nonnormal distributions. *Proceedings of the Statistical Computing Section and Section on Statistical Graphics, American Statistical Association,* Washington DC, 52-57.

Headrick, T. C., & Sawilowsky, S. S. (2000). Properties of the rank transformation in factorial analysis of covariance. *Communications in Statistics*: *Simulation and Computation*, *29*, 1059-1088.

Headrick, T. C., & Sawilowsky, S. S. (2000). Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics*, *25*, 417-436.

Headrick, T. C., & Rotou, O. (2001). An investigation of the rank transformation in multiple regression. *Computational Statistics and Data Analysis*, in press.

Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika*, *64*, 25-35.

Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.

Iman, R. L., & Conover, W. J. (1979). The use of rank transform in regression. *Technometrics*, *21*, 499-509.

Lahey, Computer Systems, Inc. (1994). *Personal Fortran (version 3.0)*. Incline Village, NV: Author.

Puri, M. L., & Sen, P. K. (1969a). Analysis of covariance based on general rank scores. *Annals of Mathematical Statistics*, *40*, 610-618.

Salter, K. C., & Fawcett, R. F. (1993). The ART test of interaction: A robust and powerful rank test of interaction in factorial models. *Communications in Statistics*: *Computation and Simulation*, *22*, 137-153.

Sawilowsky, S. S. (1990). Non-parametric tests of interaction in experimental design. *Review of Educational Research*, *60*, 91-126.

Thompson, G. L. (1991). A note on the rank transform for interactions. *Biometrika*, *78*, 697-701.

Thompson, G. L. (1993). A correction note on the rank transform for interactions. *Biometrika*, *80*, 711.

Toothaker, L. E., & Newman, D. (1994). Nonparametric competitors in the two-way ANOVA. *Journal of Educational and Behavioral Statistics*, *19*, 237-273.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental designs* (3rd). New York: McGraw-Hill.

Wolfram, S. (1999). *The Mathematica book*, (4th). Cambridge, UK: Wolfram Media-Cambridge University Press.

Send correspondence to: Todd C. Headrick, Measurement and Statistics, Southern Illinois University at Carbondale, 227-J Wham Building, Carbondale, IL 62901-4618. Email: headrick@siu.edu.