Journal of Modern Applied Statistical Methods

Volume 6 | Issue 1 Article 6

5-1-2007

# Application of a New Procedure for Power Analysis and Comparison of the Adjusted Univariate and Multivariate Tests in Repeated Measures Designs

Sean W. Mulvenon
*University of Arkansas*

M. Austin Betz
*Arizona State University*

Kening Wang
*University of Arkansas*

Bruno D. Zumbo
*University of British Columbia*, bruno.zumbo@ubc.ca

Journal of Modern Applied Statistical Methods
May, 2007, Vol. 6, No. 1, 36-52

1538 – 9472/07/$95.00

*Regular Articles*

# Application of a New Procedure for Power Analysis and Comparison of the Adjusted Univariate and Multivariate Tests in Repeated Measures Designs

Sean W. Mulvenon
University of Arkansas

M. Austin Betz
Arizona State University

Kening Wang
University of Arkansas

Bruno Zumbo
University of British Columbia

A relationship between the multivariate and univariate noncentrality parameters in repeated measures designs was developed for the purpose of assessing the relative power of the univariate and multivariate approaches. An application is provided examining the use of repeated measures designs to evaluate student achievement in a K-12 school system.

Key words: Repeated measures designs, adjusted degrees of freedom test, noncentrality parameter, sphericity.

## Introduction

Repeated measures designs are used frequently by social and behavioral science researchers (Maxwell & Delaney, 2004; Keselman, H. J., Huberty, Lix, Olejnik, Cribbie, Donahue, Kowalchuk, Lowman, Petoskey, Keselman, J. C., & Levin, 1998). A major advantage of repeated measures designs is that subjects serve as their own controls, thus variability among the subjects due to individual differences is removed, and test results are more powerful. Various procedures can be used to do variance analysis in repeated measures designs. In addition to the traditional approaches, univariate and multivariate analyses, some methods such as Improved General Approximate method (Huynh, 1978), multivariate Welch (1951)/James (1951)-type test (*WJ* test), mixed model approach (Littell, Milliken, Stroup, & Wolfinger, 1996), and empirical Bayes method (Boik, 1997) have also been studied and recommended. Guidelines for choosing an analysis strategy are generally based on whether the design is balanced or not (Keselman, 1998; Keselman, Algina, & Kowalchuk, 2002). If group sizes are equal and there is no missing data, univariate and multivariate methods are frequently used by researchers and are recommended as appropriate statistical methods (Kirk, 1995; Morrison, 1990; Maxwell & Delaney, 2004).

Both of the univariate and multivariate methods require the data satisfy certain assumptions: independent observations, multivariate normality, and homogeneous variance/covariance across groups. In addition to the above assumptions, the univariate analysis has the additional assumption of sphericity (Huynh & Feldt, 1970; Rouanet & Lépine, 1970). Sphericity refers to differences between any pair of repeated measures are equally variable. If sphericity is met, the univariate analysis has greater power than the multivariate analysis (due to a clear degrees of freedom advantage in the denominator), and it allows the

Sean W. Mulvenon is Professor of Educational Statistics and the Billingsley Chair for Educational Research and Policy Studies. Email him at seanm@uark.edu. M. Austin Betz is Professor Emeritus at Arizona Statue University. Kening Wang is a Research Associate in statistics and SAS programmer at the National Office for Research on Measurement and Evaluation Systems. Bruno D. Zumbo is a Professor of Educational Statistics at the University of British Columbia and a research fellow at the Assessment and Testing Research Unit at Carleton University. Email him at bruno.zumbo@ubc.ca

use of fewer subjects than the multivariate analysis for equal power (Morrison, 1990). Unfortunately, the assumption of sphericity is not often met in the behavioral and social research (Davidson, 1972; McCall & Appelbaum, 1973; Rogan, Keselman, & Mendoza, 1979; Keselman, Huberty *et al.*, 1998). If sphericity is not satisfied, the univariate analysis produces biased tests of significance (Box, 1954), and an adjusted degrees of freedom test, such as Greenhouse & Geisser (1959) or Huynh & Feldt (1976) test is suggested. The adjusted univariate analyses modify the *df* of the traditional *F* statistic using a sample estimate of the sphericity parameter epsilon ($\in$). The $\in$ is a measure of the degree of violation of the sphericity assumption, with perfect conformity to sphericity producing a $\in$ of 1.0 (Huynh & Feldt, 1970).

Because the empirical evidence indicates that if the design is balanced, both the adjusted univariate and the multivariate approaches give the necessary control of Type I error (Davidson, 1972; Maxwell & Arvey, 1982; Muller & Barton, 1989; Keselman, J., Lix, & Keselman, H., 1996), power becomes a critical factor in the selection between the adjusted univariate analysis and the multivariate analysis. Prospective power analysis will help researchers to determine an appropriate sample size to obtain the desired level of power to detect the meaningful differences that are hypothesized. Selecting an insufficient sample size will increase the risk of failing to detect an important difference when it may exist (Type II error). Conversely, selecting an excessive sample size may produce a statistically significant result, but one with limited meaningfulness due to small differences.

Sample size also affects the relative power of the adjusted univariate and multivariate tests. Without sphericity, multivariate tests may be more powerful than the adjusted univariate tests (Davidson, 1972). However, if sample sizes are small, the adjusted univariate analysis may still be more powerful than the multivariate analysis, because the estimators of the covariance parameters lack precision, and as a result, the power of the multivariate analysis is low (Boik, 1981). As sample size increases, the power of the multivariate test improves and can be greater than the power of the adjusted univariate test (Boik, 1997).

Power analysis and minimum sample size calculations are needed for choosing the most suitable method under different conditions. Using the expansions of Fujikoshi (1973), Sugiura (1973), or Vander Merwe and Crowther (1984), power of the multivariate tests can be computed. Muller and Peterson (1984) provided power approximations of the multivariate tests. For the adjusted univariate tests, Muller and Barton (1989, 1991) provided power approximations based on the expected value approximations for the epsilon ($\in$) estimator. Vonesh and Schork (1986) presented a statistical methodology for determining the minimum sample size for the within-subjects repeated-measures design.

They developed a formulae for calculating the multivariate noncentrality parameter, subject to constraint $\Delta = |\mu_j - \mu_k|$, which represents a minimal difference between any pair of treatment means. Rochon (1991) extend the procedures of Vonesh and Schork to the between-subjects repeated-measures design when there are only two treatment groups under consideration. All of the above researches provide strong basis for the purpose of the current paper, that is, to develop a relationship between the multivariate and univariate noncentrality parameters for assessing the relative power of the univariate and multivariate approaches in repeated measures designs. A major goal of this article is to compare the statistical power of the univariate and multivariate procedures and provide a method for selecting an appropriate sample size, given a desired effect size and level of power, when researchers are developing a study.

## Theoretical Foundations and Statistics

### The Model and Hypothesis

The usual general linear model with *g* between-subject groups and one within-subject repeated measures factor having *p* levels can be written as follows:

$$Y = XM + E \tag{1}$$

Where Y is an $N\times p$ matrix from $N$ subjects; X is an $N \times g$ between-subject design matrix; M is an $g \times p$ parameters matrix; and E is an $N \times p$ matrix of random errors. The rows of E are assumed to be independently and identically distributed as $N_p$ (0, Σ), where Σ is a $p \times p$ positive definite covariance matrix.

The general linear multivariate model hypothesis has the usual form:

$$H_0 : \Theta = C\,M\,U = \Theta_0 \tag{2}$$

where $C$ is an $a \times g$ between-subject contrasts; $M$ is an $g \times p$ parameters matrix; $U$ is an $p \times b$ within-subject contrasts; and $\Theta$ is an $a \times b$ secondary parameters matrix. Without loss of generality, assume $\Theta_0 = 0$. Define $\Sigma_* = U'\Sigma U$, which is a covariance matrix of $\text{row}_i\,(EU)'$. Also define $\omega = (\Theta-\Theta_0)'[C(X'X)^{-1}C']^{-1}\ (\Theta-\Theta_0)$, which is an unscaled noncentrality matrix. Then, the scaled noncentrality matrix (Ω) and its trace ($\delta_m$) are given by $\Omega = \omega\Sigma^{-1}_*$ and $\delta_m = \text{tr}\ (\Omega)$ respectively. Using two theorems (Theorem 2, p30; and Theorem 3, p31) from Magnus and Neudecker (1988), the general form of the noncentrality parameter for the F-distribution can be written as:

$$\delta_m = (\text{vec}M')'(C'\otimes U)\left[\left[C(X'X)^{-1}C'\right]^{-1}\otimes\Sigma_*^{-1}\right](C\otimes U')(\text{vec}M') \tag{3}$$

The hypothesis in (2) can be tested using the multivariate test. If using the Hotelling-Lawley trace statistic, the noncentrality parameter has the form (Muller & Peterson, 1984; Muller, LaVange, Ramey, & Ramey, 1992):

$$\delta_{HLT} = (ab)\cdot F_A(HLT) = \frac{HLT_A/S}{1/df_d(HLT)} \tag{4}$$

where $a = g - 1$, $b = p - 1$, $S = \min(a, b)$, and $df_d\,(HLT) = S\,[(N - g) - b - 1] + 2$.

The hypotheses in (2) also can be tested using the adjusted univariate test. Additionally, for repeated measures designs, the univariate analysis is a simply by-product of the multivariate analysis (Wang, 1983). A Univariate noncentrality parameter can be derived from (3) and be expressed as:

$$\delta_u = b\varepsilon\frac{tr(\omega)}{tr(\Sigma_*)} = \left(\frac{tr(\Sigma_*)}{tr(\Sigma_*^2)}\right)\cdot tr(\omega) \tag{5}$$

where

$$\varepsilon = \frac{tr^2\left(\Sigma_*\right)}{b\cdot tr\left(\Sigma_*^2\right)}\ \text{(Box, 1954).}$$

The sphericity parameter $\epsilon$ ($1/b \le \epsilon \le 1$) reflects a discrepancy from sphericity. If sphericity is not met ($\epsilon \ne 1$), $T^2_{\text{univariate}} \approx F\ (ab\epsilon,\ b(N\text{-}g)\epsilon,\ \delta_u)$. The Greenhouse & Geisser test uses the maximum likelihood estimator (MLE) $\hat{\varepsilon}$ to adjust the degrees of freedom of the univariate test.

Lower Bounds of Noncentrality Parameters

The noncentral $F$-distribution can be used for power and sample size calculations. The power associated with the $F$-test is a monotonically increasing function of the noncentrality parameter. In this subsection, minimizing of the noncentrality parameters, developing the lower bounds of the multivariate and univariate noncentrality parameters using the same constraints, and establishing a relationship between them are described.

As shown and demonstrated in the Appendix, for fixed $\epsilon > 0$, subject to the constraint

$$\Delta = \frac{|c'Md|}{\sqrt{c'c}\sqrt{d'd}} \quad\text{and}\quad \bar{\theta} = \frac{tr(\Sigma_*)}{b} \tag{6}$$

the lower bounds of multivariate and univariate noncentrality parameters can be expressed as:

$$\delta_m^* = \frac{N\Delta^2}{d'\Sigma d} \quad\text{and}\quad \delta_u^* = \frac{\varepsilon N\Delta^2}{\bar{\theta}} \tag{7}$$

where $c$ and $d$ are arbitrary vectors of contrast coefficients. $\delta_m \geq \delta^*_m$, and $\delta_u \geq \delta^*_u$. The value of $\overline{\theta}$ is represented as:

$$\overline{\theta} = \frac{tr(\Sigma_*)}{b} = \frac{\sum_{i=1}^{b} \lambda_i}{b} \qquad (8)$$

which is the mean eigenvalue of the error matrix $\Sigma_*$.

From (7) and (8), the relationship between the lower bounds of multivariate and univariate noncentrality parameters can be expressed as:

$$\delta^*_u = \varepsilon^* \delta^*_m \qquad (9)$$

where $\varepsilon^* = \varepsilon\phi$, $\phi = \frac{\sigma_0^2}{\overline{\theta}}$, and $\phi$ represents the bias ratio.

As shown by Boik (1981), $\sigma_0^2 = d'\Sigma d$ is the experimental error of contrast among the $p$-repeated measures, and $\overline{\theta}$ is the average experimental error of any set of $b = p - 1$ orthonormal contrasts. Further, when the sphericity assumption is met ($\epsilon = 1.0$), $\phi$ will always equal unity (Boik, 1981). For fixed $\epsilon$, the bias ratio has a range of values, $\phi_{\min} < \phi < \phi_{\max}$. The upper and lower limits of $\phi$ are given by Boik (1981):

$$\phi_{\max} = 1 + B, \textit{ when } 1/b \leq \epsilon \leq 1,$$

$$\phi_{\min} = \begin{cases} 1 - B & \textit{when} \quad \frac{b-1}{b} \leq \varepsilon \leq 1.0 \\ 0 & \textit{when} \quad \frac{1}{b} \leq \varepsilon \leq \frac{b-1}{b} \end{cases} \qquad (10)$$

where $B = \left[\frac{(b-1)(1-\varepsilon)}{\varepsilon}\right]^{1/2}$. If $\epsilon = 1$, then $B = 0$, and $\phi_{\min} = \phi_{\max} = 1$.

Because $\varepsilon^* = \varepsilon\phi$ and $\phi_{\min} < \phi < \phi_{\max}$, the maximum and minimum values of multiplier $\varepsilon^*$ can be obtained as:

$$\varepsilon^*_{\max} = \varepsilon(1+B) = \varepsilon + \sqrt{(b-1)(1-\varepsilon)\varepsilon} \qquad (11)$$

and

$$\varepsilon^*_{\min} = \varepsilon(1-B) = \varepsilon - \sqrt{(b-1)(1-\varepsilon)\varepsilon} \qquad (12)$$

$\varepsilon^*_{\max}$ varies between a minimum of 1, when $\epsilon = 1/b$ or $\epsilon = 1$; and a maximum of $\frac{1}{2}(1+\sqrt{b})$, when $\varepsilon = \frac{1}{2}(1+\sqrt{\frac{1}{b}})$. $\varepsilon^*_{\min}$ varies between a minimum of 0, when $\epsilon = (b-1)/b$; and a maximum of 1, when $\epsilon = 1$. For example, let $\varepsilon = \frac{1}{2}(1+\sqrt{\frac{1}{b}})$ and $(b-1)/b < \epsilon \leq 1$, then $\varepsilon^*_{\max} = \frac{1}{2}(1+\sqrt{b})$, and $\varepsilon^*_{\min} = \frac{1}{2}\left[1-(b-2)\sqrt{\frac{1}{b}}\right]$. Let $\varepsilon = \frac{1}{2}(1+\sqrt{\frac{1}{b}})$, but if $\epsilon$ is in the interval of $[1/b, (b-1)/b]$, then $\varepsilon^*_{\max} = \frac{1}{2}(1+\sqrt{b})$, but $\varepsilon^*_{\min} = 0$, because of the restrictive nature of the bias ratio $\phi_{\min} = 0$, when $1/b \leq \epsilon \leq (b-1)/b$.

Best and Worst Case Scenarios for the Univariate Test

An examination of (9), (11), and (12) allows the determination of best and worst case scenarios for the lower bound of univariate noncentrality parameter ($\delta^*_u$) by substituting the maximum and minimum values of $\varepsilon^*$ in (9). The best case scenario for $\delta^*_u$ is

$$\delta_u^*\_best = \left[\varepsilon + \sqrt{(b-1)(1-\varepsilon)\varepsilon}\right] \cdot \delta_m^* \tag{13}$$

Because

$$1 \le \varepsilon_{\max}^* = \left[\varepsilon + \sqrt{(b-1)(1-\varepsilon)\varepsilon}\right] \le \frac{1}{2}(1+\sqrt{b})$$

this suggests the best case scenario for the univariate case, which means the minimum power of the univariate test will generally exceed the minimum power of the multivariate test.

However, substituting $\varepsilon_{\min}^*$ in (9) yields the worst case scenario:

$$\delta_u^*\_worst = \left[\varepsilon - \sqrt{(b-1)(1-\varepsilon)\varepsilon}\right] \cdot \delta_m^* \tag{14}$$

Because

$$0 \le \varepsilon_{\min}^* = \left[\varepsilon - \sqrt{(b-1)(1-\varepsilon)\varepsilon}\right] \le 1,$$

this suggests the worst case scenario for the univariate case, which means the minimum power of the univariate test will be generally lower than the minimum power of the multivariate test.

Univariate versus Multivariate

Power Analysis and Minimum Sample Size Calculation

For computing the minimum necessary sample size to obtain a desired level of power in the multivariate case, Vonesh and Schork (1986) presented a statistic method, and Rochon (1991) extended it to the between-subjects repeated-measures design. If let $\Sigma$ to be a positive covariance matrix, which means $\rho_{jk} \ge 0$ for $j < k$; and let $\sigma_{\max}^2$ represents the largest variance, then the lower bound of $\delta_m$ can be approximated:

$$\delta_m^* = \frac{N\Delta^2}{2\sigma_{\max}^2(1-\rho_{\min})} \le \delta_m \tag{15}$$

where $\rho_{\min} = \min_t \{\rho_t\}$. This would guarantee power greater than the normal level. Using the above approximation, the minimum sample size for the multivariate case can be determined by utilizing

$$1-\beta_m = P\left[F\left(df_n, df_d; \delta_m\right) > F_\alpha\left(df_n, df_d\right)\right] \tag{16}$$

To determine sample sizes in the univariate case when the assumption of sphericity is untenable, the following is used

$$1-\beta_u = P\begin{bmatrix} F(ab\varepsilon, b(N-g)\varepsilon; \delta_u) \\ > F_\alpha\left(ab\varepsilon, b(N-g)\varepsilon\right) \end{bmatrix} \tag{17}$$

where $\delta_u$ and $\epsilon$ are given in (5).

In order to determine the minimum sample sizes in the univariate case, applying (13) and (14), the upper (best case) and lower (worst case) limits of the $\delta_u^*$ can be obtained, if $\epsilon$ and $\delta_m^*$ are known. $\delta_m^*$ can be approximated by (15). In general, however, it will not be known. In the present context, suppose $\varepsilon = \frac{1}{2}(1+\sqrt{\frac{1}{b}})$, then if $\epsilon$ is in the interval $[(b-1)/b, 1]$, the upper (best case scenario) limit of the $\delta_u^*$ can be obtained as $\delta_u^* = \frac{1}{2}(1+\sqrt{b})\delta_m^*$; and the lower (worst case) limit of the $\delta_u^*$ can be obtained as $\delta_u^* = \frac{1}{2}\left[1-(b-2)\sqrt{\frac{1}{b}}\right]\delta_m^*$. This enables determination of the upper and lower limits of the $\delta_u^*$ for simulation study of the best and the worst case scenarios for the univariate case.

Simulation Procedure

The simulation was conducted in SAS/IML and SAS program is available from the author on request. The process of minimum sample size determination, or statistical power analysis, involves the following four components: Type I error ($\alpha$), power ($1-\beta$), effect size $\Delta$ or standardized effect size $\Delta_*$, and the minimum correlation $\rho_{min}$. Desired statistical power is set to be 0.80 in this study. The 80% level of power is based on Cohen's well-informed conjecture that the rate of Type II error should be about fourfold that of Type I error (Cohen, 1992). Detailed procedures were given as the following steps:

1) Specify the desired power ($1 - \beta$) to be 0.8, and $\alpha = 0.05$. Set all possible combinations of the following values: $p = 3, 4$; $g = 2, 3, 4$; $\rho_{min} = .1, .2, \ldots, .9$ by .1; and $\Delta_* = .2, .3, \ldots, 1.5$ by .1.

2) The necessary sample size ($N_m$) was computed for all the above combinations using the multivariate procedure.

3) Using the upper limit of the $\delta_u^*$ ($\delta_u^* = \frac{1}{2}(1+\sqrt{b})\delta_m^*$) to calculate the necessary sample size ($N_u$) for the best case scenario of the univariate procedure.

4) Using the lower limit of the $\delta_u^*$ ($\delta_u^* = \frac{1}{2}\left[1-(b-2)\sqrt{\frac{1}{b}}\right]\delta_m^*$) to calculate the necessary sample size ($N_l$) for the worst case scenario of the univariate procedure.

Monte Carlo

Table 1 contains a selection of the results from the univariate and multivariate simulations. A comparison of the minimum sample size estimates between the multivariate procedure and the univariate procedure for the best and the worst case scenarios indicates some clear trends.

First, when the effect size is small, for example, $\Delta_* \leq 0.4$, and if minimum correlation is also small, then the minimum sample sizes of the multivariate procedure ($N_m$) are much larger than the univariate procedure for the best case scenario ($N_u$). This trend indicates that when the above conditions hold, researchers need to consider using the univariate procedure, especially when sample sizes anticipated for the study may be small. This result is consistent with Boik's (1997) conclusion that if sample sizes are small, the adjusted univariate analysis may still be more powerful than the multivariate analysis. When the design becomes more complex, this trend is more obvious, because the minimum sample sizes generally increase as the number of groups and repeated trials increases (due to space considerations, results of other combinations of groups and trials are not included in the table).

Second, when the effect size is large, for example, $\Delta_* \geq 0.8$, the multivariate procedure could generally be recommended due to small minimum sample sizes. Simulation results indicate that there is small degree of divergence of the minimum sample sizes between the multivariate procedure ($N_m$) and the univariate procedure for the best case scenario ($N_u$).

Third, when the effect size is moderate, for example, $0.4 < \Delta_* < 0.8$, the minimum correlation ($\rho_{min}$) will provide valuable information in selecting between the univariate and multivariate procedures. If $\rho_{min}$ is large, for example, $\rho_{min} \geq .80$, then the univariate procedure is recommended; otherwise, researchers need to consider using the multivariate procedure.

Upon inspection of this table, a pattern was also found for the relationship between the minimum sample size and the effect size. For fixed power, the minimum sample size generally decreases as the effect size increases. Thus, if sample size is fixed, larger treatment differences will provide greater power. The same pattern can be observed for the relationship between the minimum sample size and the minimum correlation.

Table 1. Necessary sample size estimates by groups, trials, standardized effect size ($\Delta_*$), and minimum correlation ($\rho_{min}$) for desired power = .80 at $\alpha$ = .05

| Groups | trials | $\Delta_*$ | $\rho_{min}$ | $N_m$ | $N_u$ | $N_l$ |
|---|---|---|---|---|---|---|
| 2 | 3 | 0.2 | 0.1 | 227 | 9 | > 500 |
| 2 | 3 | 0.2 | 0.2 | 202 | 9 | > 500 |
| 2 | 3 | 0.2 | 0.3 | 177 | 9 | > 500 |
| 2 | 3 | 0.2 | 0.4 | 152 | 9 | > 500 |
| 2 | 3 | 0.2 | 0.5 | 127 | 9 | > 500 |
| 2 | 3 | 0.2 | 0.6 | 103 | 8 | > 500 |
| 2 | 3 | 0.2 | 0.7 | 78 | 8 | > 500 |
| 2 | 3 | 0.2 | 0.8 | 53 | 8 | > 500 |
| 2 | 3 | 0.2 | 0.9 | 28 | 7 | > 500 |
| 2 | 3 | 0.3 | 0.1 | 103 | 8 | > 500 |
| 2 | 3 | 0.3 | 0.2 | 92 | 8 | > 500 |
| 2 | 3 | 0.3 | 0.3 | 81 | 8 | > 500 |
| 2 | 3 | 0.3 | 0.4 | 70 | 8 | > 500 |
| 2 | 3 | 0.3 | 0.5 | 59 | 8 | > 500 |
| 2 | 3 | 0.3 | 0.6 | 48 | 7 | > 500 |
| 2 | 3 | 0.3 | 0.7 | 37 | 7 | > 500 |
| 2 | 3 | 0.3 | 0.8 | 26 | 7 | > 500 |
| 2 | 3 | 0.3 | 0.9 | 15 | 6 | > 500 |
| 2 | 3 | 0.4 | 0.1 | 59 | 8 | > 500 |
| 2 | 3 | 0.4 | 0.2 | 53 | 8 | > 500 |
| 2 | 3 | 0.4 | 0.3 | 47 | 8 | > 500 |
| 2 | 3 | 0.4 | 0.4 | 41 | 7 | > 500 |
| 2 | 3 | 0.4 | 0.5 | 35 | 7 | > 500 |
| 2 | 3 | 0.4 | 0.6 | 28 | 7 | > 500 |
| 2 | 3 | 0.4 | 0.7 | 22 | 6 | > 500 |
| 2 | 3 | 0.4 | 0.8 | 16 | 6 | > 500 |
| 2 | 3 | 0.4 | 0.9 | 10 | 5 | > 500 |
| 2 | 3 | 0.5 | 0.1 | 39 | 7 | > 500 |
| 2 | 3 | 0.5 | 0.2 | 35 | 7 | > 500 |
| 2 | 3 | 0.5 | 0.3 | 31 | 7 | > 500 |
| 2 | 3 | 0.5 | 0.4 | 27 | 7 | > 500 |
| 2 | 3 | 0.5 | 0.5 | 24 | 6 | > 500 |
| 2 | 3 | 0.5 | 0.6 | 20 | 6 | > 500 |
| 2 | 3 | 0.5 | 0.7 | 16 | 6 | > 500 |
| 2 | 3 | 0.5 | 0.8 | 12 | 5 | > 500 |
| 2 | 3 | 0.5 | 0.9 | 8 | 5 | 26 |
| 2 | 3 | 0.6 | 0.1 | 28 | 7 | > 500 |
| 2 | 3 | 0.6 | 0.2 | 26 | 7 | > 500 |
| 2 | 3 | 0.6 | 0.3 | 23 | 6 | > 500 |
| 2 | 3 | 0.6 | 0.4 | 20 | 6 | > 500 |
| 2 | 3 | 0.6 | 0.5 | 18 | 6 | > 500 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | 3 | 0.6 | 0.6 | 15 | 6 | > 500 |
| 2 | 3 | 0.6 | 0.7 | 12 | 5 | > 500 |
| 2 | 3 | 0.6 | 0.8 | 10 | 5 | > 500 |
| 2 | 3 | 0.6 | 0.9 | 7 | 5 | 8 |
| 2 | 3 | 0.7 | 0.1 | 22 | 6 | > 500 |
| 2 | 3 | 0.7 | 0.2 | 20 | 6 | > 500 |
| 2 | 3 | 0.7 | 0.3 | 18 | 6 | > 500 |
| 2 | 3 | 0.7 | 0.4 | 16 | 6 | > 500 |
| 2 | 3 | 0.7 | 0.5 | 14 | 5 | > 500 |
| 2 | 3 | 0.7 | 0.6 | 12 | 5 | > 500 |
| 2 | 3 | 0.7 | 0.7 | 10 | 5 | > 500 |
| 2 | 3 | 0.7 | 0.8 | 8 | 5 | 37 |
| 2 | 3 | 0.7 | 0.9 | 6 | 5 | 6 |
| 2 | 3 | 0.8 | 0.1 | 18 | 6 | > 500 |
| 2 | 3 | 0.8 | 0.2 | 16 | 6 | > 500 |
| 2 | 3 | 0.8 | 0.3 | 15 | 5 | > 500 |
| 2 | 3 | 0.8 | 0.4 | 13 | 5 | > 500 |
| 2 | 3 | 0.8 | 0.5 | 12 | 5 | > 500 |
| 2 | 3 | 0.8 | 0.6 | 10 | 5 | > 500 |
| 2 | 3 | 0.8 | 0.7 | 9 | 5 | 63 |
| 2 | 3 | 0.8 | 0.8 | 7 | 5 | 11 |
| 2 | 3 | 0.8 | 0.9 | 6 | 5 | 5 |
| 2 | 3 | 0.9 | 0.1 | 15 | 6 | > 500 |
| 2 | 3 | 0.9 | 0.2 | 14 | 5 | > 500 |
| 2 | 3 | 0.9 | 0.3 | 13 | 5 | > 500 |
| 2 | 3 | 0.9 | 0.4 | 11 | 5 | > 500 |
| 2 | 3 | 0.9 | 0.5 | 10 | 5 | > 500 |
| 2 | 3 | 0.9 | 0.6 | 9 | 5 | > 500 |
| 2 | 3 | 0.9 | 0.7 | 8 | 5 | 14 |
| 2 | 3 | 0.9 | 0.8 | 7 | 5 | 6 |
| 2 | 3 | 0.9 | 0.9 | 6 | 5 | 5 |
| 2 | 3 | 1.0 | 0.1 | 13 | 5 | > 500 |
| 2 | 3 | 1.0 | 0.2 | 12 | 5 | > 500 |
| 2 | 3 | 1.0 | 0.3 | 11 | 5 | > 500 |
| 2 | 3 | 1.0 | 0.4 | 10 | 5 | > 500 |
| 2 | 3 | 1.0 | 0.5 | 9 | 5 | > 500 |
| 2 | 3 | 1.0 | 0.6 | 8 | 5 | 26 |
| 2 | 3 | 1.0 | 0.7 | 7 | 5 | 10 |
| 2 | 3 | 1.0 | 0.8 | 6 | 5 | 6 |
| 2 | 3 | 1.0 | 0.9 | 6 | 5 | 5 |
| 2 | 3 | 1.1 | 0.1 | 11 | 5 | > 500 |
| 2 | 3 | 1.1 | 0.2 | 11 | 5 | > 500 |
| 2 | 3 | 1.1 | 0.3 | 10 | 5 | > 500 |
| 2 | 3 | 1.1 | 0.4 | 9 | 5 | > 500 |
| 2 | 3 | 1.1 | 0.5 | 8 | 5 | 49 |
| 2 | 3 | 1.1 | 0.6 | 8 | 5 | 9 |
| 2 | 3 | 1.1 | 0.7 | 7 | 5 | 7 |
| 2 | 3 | 1.1 | 0.8 | 6 | 5 | 5 |
| 2 | 3 | 1.1 | 0.9 | 5 | 5 | 5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | 3 | 1.2 | 0.1 | 10 | 5 | > 500 |
| 2 | 3 | 1.2 | 0.2 | 10 | 5 | > 500 |
| 2 | 3 | 1.2 | 0.3 | 9 | 5 | > 500 |
| 2 | 3 | 1.2 | 0.4 | 8 | 5 | 63 |
| 2 | 3 | 1.2 | 0.5 | 8 | 5 | 10 |
| 2 | 3 | 1.2 | 0.6 | 7 | 5 | 8 |
| 2 | 3 | 1.2 | 0.7 | 6 | 5 | 6 |
| 2 | 3 | 1.2 | 0.8 | 6 | 5 | 5 |
| 2 | 3 | 1.2 | 0.9 | 5 | 5 | 5 |
| 2 | 3 | 1.3 | 0.1 | 9 | 5 | > 500 |
| 2 | 3 | 1.3 | 0.2 | 9 | 5 | 100 |
| 2 | 3 | 1.3 | 0.3 | 8 | 5 | 52 |
| 2 | 3 | 1.3 | 0.4 | 8 | 5 | 11 |
| 2 | 3 | 1.3 | 0.5 | 7 | 5 | 9 |
| 2 | 3 | 1.3 | 0.6 | 7 | 5 | 6 |
| 2 | 3 | 1.3 | 0.7 | 6 | 5 | 5 |
| 2 | 3 | 1.3 | 0.8 | 6 | 5 | 5 |
| 2 | 3 | 1.3 | 0.9 | 5 | 5 | 5 |
| 2 | 3 | 1.4 | 0.1 | 9 | 5 | 37 |
| 2 | 3 | 1.4 | 0.2 | 8 | 5 | 37 |
| 2 | 3 | 1.4 | 0.3 | 8 | 5 | 11 |
| 2 | 3 | 1.4 | 0.4 | 7 | 5 | 11 |
| 2 | 3 | 1.4 | 0.5 | 7 | 5 | 7 |
| 2 | 3 | 1.4 | 0.6 | 6 | 5 | 6 |
| 2 | 3 | 1.4 | 0.7 | 6 | 5 | 5 |
| 2 | 3 | 1.4 | 0.8 | 6 | 5 | 5 |
| 2 | 3 | 1.4 | 0.9 | 5 | 5 | 5 |
| 2 | 3 | 1.5 | 0.1 | 8 | 5 | 26 |
| 2 | 3 | 1.5 | 0.2 | 8 | 5 | 11 |
| 2 | 3 | 1.5 | 0.3 | 7 | 5 | 11 |
| 2 | 3 | 1.5 | 0.4 | 7 | 5 | 7 |
| 2 | 3 | 1.5 | 0.5 | 7 | 5 | 6 |
| 2 | 3 | 1.5 | 0.6 | 6 | 5 | 5 |
| 2 | 3 | 1.5 | 0.7 | 6 | 5 | 5 |
| 2 | 3 | 1.5 | 0.8 | 5 | 5 | 5 |
| 2 | 3 | 1.5 | 0.9 | 5 | 5 | 5 |

*Notes:* [*] Due to space considerations, not all of the simulation results are included in the table, but they are available from the author on request.

[1] $N_m$ represents the necessary sample size computed using the multivariate procedure.

[2] $N_u$ represents the necessary sample size computed for the "best case scenario" of the univariate procedure.

[3] $N_l$ represents the necessary sample size computed for the "worst case scenario" of the univariate procedure.

A Case Study: Examination of Student Achievement Models

The most effective method to evaluate student achievement is to monitor change in performance between two or more points, or more specifically a repeated measures design. Recent "No Child Left Behind" (NCLB) legislation has contributed to a proliferation of growth models advocated as best methods to examine student achievement. A major concern with the use of most of these growth models is they assume large samples. However, within most traditional educational settings, sample sizes are relatively small. The use of the more traditional repeated measures designs, univariate or multivariate, may be more appropriate than hierarchical linear models or latent growth analyses.

Case Study

A recent and growing concern in K-12 education has been the preparation of students to be successful in college. To address this issue, numerous studies have been completed that examine a student's high school record of achievement. However, education is a linear system, with students in theory, starting at grade one and progressing through the system to grade twelve. Additionally, in large school districts, a significant amount of concern is directed at the preparation of student's prior to high school. This case study examines three elementary schools and the difference in performance of students as they progress through this K-12 school system.

Each elementary school has grades kindergarten through fifth grade. Students were administered standardized reading tests in fifth, seventh (while at a middle school within the same district), and tenth grade. The primary research question, does elementary school you attended makes a difference in determining your starting point ($10^{th}$ grade) at the local high school? Table 2 provides a means and standard deviations of scaled scores for students from each of the three elementary schools. The small sample sizes reflect the issue of mobility of students, and in particular from School A, where annually 30 percent of students are identified as highly mobile.

A total of four analyses were completed: (1) School A versus School B, (2) School A versus School C, (3) School B versus School C, and (4) School A, School B, and School C. Table 3 provides the multivariate and univariate results in addition to retrospective and prospective power estimation values. The result demonstrated the importance of the univariate procedure with large effect sizes and a limited number of observations. Additionally, it is expected that standardized tests will have a strong correlation from year to year, which also contributes to the strength of the univariate procedure.

The case study was done as a study of convenience with data that represented the most common type of educational data used to complete school evaluations. In practice, analyses will be completed at the classroom, grade or school level in efforts to evaluate the impact of instructional practices or new educational interventions. The present case study does an excellent job of replicating the sample size and type of outcome variables (standardized test) that will be employed and demonstrated, in practice, why greater consideration needs to be given to use of the univariate method in repeated measures designs.

Table 2. Case Study: School Test Scores

| School | N | Score 1 | Score 2 | Score 3 |
|---|---|---|---|---|
| A | 12 | 614.3(37.6) | 653.5(34.6) | 685.2(29.5) |
| B | 27 | 666.1(35.1) | 680.6(26.1) | 713.0(27.6) |
| C | 25 | 653.4(34.3) | 678.8(27.9) | 704.8(29.7) |

Table 3. Power Results for Univariate and Multivariate Comparisons

| Comparison Schools | N | Retrospective Power Delta | GG_PWR | M_PWR | Univ_F | Mult_F | N | Prospective Power Delta | PM | PU | LU | UU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A vs. B | 39 | 2.74 | 0.88 | 0.23 | 7.51(.0015) | 5.67(.0072) | 60 | 0.5 | 0.81 | 0.81 | 0.53 | 0.90 |
| A vs. C | 37 | 1.99 | 0.60 | 0.14 | 3.95(.0291) | 2.82(.0738) | 75 | 0.5 | 0.81 | 0.81 | 0.53 | 0.90 |
| B vs. C | 52 | 1.40 | 0.36 | 0.17 | 1.95(>.05) | 1.79(>.05) | 78 | 0.5 | 0.81 | 0.81 | 0.53 | 0.90 |
| A vs. B vs. C | 64 | 2.94 | 0.82 | 0.13 | 4.31(.0037) | 3.46(0.114) | 117 | 0.5 | 0.80 | 0.80 | 0.49 | 0.89 |

N = sample size
Delta = effect size
GG_PWR = Univariate Power
M_PWR = Multivariate Power
Univ_F = Univariate F-test and alpha
Mult_F = Hotelling-Lawley Trace F-test and Alpha

N = sample Size
Delta = Effect size
PM = Prospective Multivariate Power
PU = Prospective Univariate Power
LU = Lower Bound Univariate Power
UU = Upper Bound Univariate Power

## Conclusion

The relationship between $\delta_u^*$ and $\delta_m^*$, which was developed in this study, provides a theoretical foundation for calculation of prospective power estimates for the univariate case in repeated measures designs. The relationship $\delta^*_u = \epsilon^* \delta^*_m$ can be employed to compute the univariate noncentrality parameter when the multivariate noncentrality parameter has been computed. This permits calculation of minimum sample size estimates and power analysis for the univariate procedure; and it provides a basis to address the question of which procedure to propose, univariate or multivariate, when designing a study which involves repeated measures.

Some researchers have compared the benefits of using either a multivariate or univariate procedure. Barcikowski and Robey (1984) and Stevens (2002) suggested that when conducting an exploratory analysis, both the adjusted univariate and multivariate procedures should be employed because each analysis could possibly reveal different treatment effects. O'Brien and Kaiser (1985) reported after a thorough review of the literature, under no conditions is one procedure uniformly more powerful. Results from this study indicate that generally, a researcher can use the multivariate procedure in most cases, as it does provide adequate power protection. However, the univariate procedure clearly provides greater protection under some specific conditions, indicated as best case scenarios, and therefore can be recommended for these conditions.

Maxwell and Delaney (1990) provided an empirical guideline that if the sample size ($N$) is less than $p + 10$ ($p$ representing the number of repeated trials), the univariate procedure is recommended; otherwise, if $N \geq p + 10$, the multivariate procedure is recommended. In the 2nd edition, Maxwell & Delaney (2004) modified the empirical guideline, and it is that the multivariate approach probably should be used if (1) $p \leq 4$, $\varepsilon \leq .90$, and $n \geq p + 15$, or if (2) $5 \leq p \leq 8$, $\varepsilon \leq .85$, and $n \geq p + 20$. Results from this study indicate that the suggested guideline by Maxwell and Delaney works well, but only when the effect size and the minimal correlation are large.

In closing, this study effectively validates many of the recommendations of Boik, Maxwell & Delaney, and others; additionally, it expands the window where univariate repeated measures designs should be employed.

## References

Barcikowski, R. S., & Robey, R. R. (1984). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. *The American Statistician, 38,* 148-150.

Boik, R. (1981). A priori tests in repeated measures designs: Effects of nonsphericity. *Psychological Bulletin, 86*, 1084-1089.

Boik, R. J. (1997). Analysis of repeated measures under second-stage sphericity: An empirical Bayes approach. *Journal of Educational and Behavioral Statistics, 22,*155-192.

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II: Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics, 25,* 484–498.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Davidson, M. L. (1972). Univariate versus multivariate tests in repeated-measures experiments. *Psychological Bulletin, 77,* 446-452.

Fujikoshi, Y. (1973). Asymptotic formulas for the distribution of three statistics for multivariate linear hypothesis. *Annals of the Institute of Statistical Mathematics, 25*, 423-437.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24,* 95–112.

Huynh, H. & Feldt, L. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F distributions. *Journal of the American Statistical Association, 65,* 1582-1589.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split plot designs. *Journal of Educational Statistics, 1,* 69–82.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika, 43,* 161-175.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika, 38,* 324-329.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Belmont, CA: Brooks/Cole.

Keselman, J., Lix, L., & Keselman, H. (1996). The analysis of repeated measurement: A quantitative research synthesis. *British Journal of Mathematical and Statistical Psychology, 49*, 275-298.

Keselman, H. J. (1998). Testing treatment effects in repeated measures designs: An update for psychophysiological researchers. *Psychophysiology*, 35, 470-478.

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. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of Educational Research, 68,* 350-386.

Keselman, H. J., Algina, J., Kowalchuk, R. K. (2002). A Comparison of Data Analysis Strategies for Testing Omnibus Effects in Higher-Order Repeated Measures Designs. *Multivariate Behavioral Research, 37,* 331-357.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus*. New York: Wiley.

Maxwell, S. E., & Arvey, R. D. (1982). Small sample profile analysis with many variables. *Psychological Bulletin, 92,* 778-785.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data.* Belmont, CA: Wadsworth.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

McCall, R. B. & Appelbaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development, 44,* 401-415.

Morrison, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.

Muller, K. E., & Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Computational Statistics and Data Analysis, 2,* 143-158.

Muller, K. E., & Barton, C. N. (1989). Approximate power for repeated measures ANOVA lacking sphericity. *Journal of the American Statistical Association, 84,* 549-555.

Muller, K. E., & Barton, C. N. (1991). Correction of power formula for reported measures ANOVA lacking sphericity. *Journal of the American Statistical Association, 86,* 255-256.

Muller, K. E, LaVange, L. M, Ramey, S. L, Ramey, C. T. (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association, 87*:1209 – 1226.

O'Brian, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97,* 316-333.

Rao, C. R. (1973). *Linear statistical inference and its application* (2nd ed.). New York: Wiley.

Rochon, J. (1991). Sample size calculations for two-group repeated measures experiments. *Biometrics, 47,* 1383-1393.

Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 32,* 269-286.

Rouanet, H. & Lepine, D. (1970). Comparison between treatments in a repeated measures design: ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology, 23,* 147-163.

SAS Institute Inc. (1999). *SAS Manual for Linear Models* (Version 8). SAS Institute Inc. Cary, NC.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Sugiura, N. (1973). Further asymptotic formulas for the non-null distributions of three statistics for multivariate linear hypothesis. *Annals of the Institute of Statistical Mathematics, 25*, 153-163.

Van der Merwe, L., & Crowther, N. A. S. (1984). An approximation to the distribution of Hotelling's $T^2$-statistic. *South African Statistical Journal, 18,* 68-90.

Vonesh, E. F., & Schork, M. A. (1986). Sample sizes in the multivariate analysis of repeated measurements. *Biometrics, 42*, 601-610.

Wang, C. M. (1983). On the analysis of multivariate repeated measures designs. *Communications in Statistics Part A- Theory and Methods, 12*, 1647-1659.

Welch, B. (1951). On the comparison of several mean values: An alternative approach. B*iometrika, 38*, 330-336.

Appendix A

Proof of Rationale for Lower Bounds of Noncentrality Parameters

In (3), for matrix $C\,(g-1\times g)$, $C\,C' = I_{g-1}$ ; and for matrix $U\,(p-1\times p)$, $U\,U' = I_{p-1}$. Define vectors of contrast coefficients $c$ ($g \times 1$) and $d$ ($p \times 1$) as $a'C = c$ and $b'U = d$, where $a$ is a vector ($g-1\times 1$) and $b$ is a vector ($p-1\times 1$). Thus, $a'$ C M U b = $c'$ M d = Δ. Because Δ is a scaler, it can be expressed as the form: $b'(U'M'C')a = \Delta$. Using the vec operator, we obtain:

$$vec\left[b'\left(U'M'C'\right)a\right] = \left(a'C \otimes b'U'\right)vecM'$$

Applying the constraints in (6), and using (1f.1.3) of Rao (1973, p. 60) to (3), the lower bound of $\delta_m$ is obtained by evaluating:

$$\min_{a,b}\left\{\inf_{\left[a'C\otimes b'U'\right]vecM'} \delta_m\right\} = \frac{n\Delta^2}{d'\Sigma d} = \delta_m^*$$

For the lower bound of $\delta_u$, subject to the same constraint as used in the multivariate case, the minimum of tr(ω) is $n\Delta^2$, then replacing tr(ω) with $n\Delta^2$, the lower bound of $\delta_u$ is obtained as:

$$\delta_u^* = b\varepsilon \cdot \frac{n\Delta^2}{tr(\Sigma_*)} = \frac{\varepsilon n\Delta^2}{\bar{\theta}},$$

where

$$\bar{\theta} = \frac{tr(\Sigma_*)}{b}.$$

## Appendix B
## SAS Programming Notation and Code for Monte Carlo Procedures

```
options ls= 121 ps= 40 nodate pageno= 1;
data temp1;
 set ade.adv_multi_data_set2;
 if gender = "M" then gender1= 1;
  if gender= "F" then gender1= 2;
 if readss1= . or readss2= . or readss3= . then
delete;
 *if leanob1 in(7203014 7203010);* 7203013);
 if leanob1 in(7203014 7203010 7203013);
   /* combinations: 7203014 with (7203010*/
run;

proc glm;
 class leanob1;
 model readss1 readss2 readss3 = leanob1;
 repeated trials 3;
 means leanob1;
run;

proc sort data= temp1; by gender1; run;

%macro powerint(pdelta=, power=, alpha= );
 proc iml;
 use work.temp1;
 read all var {leanob1} into xx3;
 gender1= unique(xx3[,1]);
 read all var {leanob1 readss1 readss2 readss3}
into xx;
 groups= unique(xx[,1]);

pdelta= &pdelta;
power= &power;
alpha= &alpha;

/*************************************/
/*    Generate Basic Values for Repeated
Measures Analysis              */
/*
*************************************/
n1= nrow(xx);
study_n= n1;
t= ncol(xx);
p= ncol(xx)-1;
x= xx[, 2:t];
b= p-1;
g= ncol(groups);
dd= ncol(leano);
a= g-1;
m= (p - 1)*ncol(groups);
sum= x(|+,|);
mean1= sum/n1;
d_Mean= mean1;
xpx= t(x)*x - t(sum)*sum/n1;
s= diag(1/sqrt(vecdiag(xpx)));
corrmat= corr(x);
covmat= xpx/(n1-1);

/*************************************/
/*   Generate Contrast Matrices for RM-Design:
Group Matrix                */
/************************************

cmatrix1= vecdiag(i(g));
cmatrix2= J(g,g-1,0);
 do h= 1 to g;
  do i= 1 to a while (i < h);
   cmatrix2[h,i]= -1;
   cmatrix2[i,i]= g - i;
  end;
 end;
cmatrix1= t(cmatrix1);
cmatrix2= t(cmatrix2);

/*************************************/
/* Generate Orthonormalized Contrast Matrices
*/
/*************************************/

u_i1= j(p,p-1,0);
do k= 1 to p;
 do l= 1 to b while (l < k);
  u_i1[k,l]= -1;
  u_i1[l,l]= p - l;
 end;
end;

u_i1=
u_i1/shape(sqrt(u_i1[##,]),nrow(u_i1),ncol(u_i1
));

/*************************************/
```

```
/*  The next piece is the iterative do-loop to make this program a generalized form. Generating */
/*  the necessary within matrix components regardless of the number of groups or subjects within a group */

do i=1 to ncol(groups);
  do rm=1 to p;
    subset= subset||remove(xx[,rm+1],loc(choose(xx[,1]=groups[i],0,1)))`;
  end;
n= nrow(subset);
    nn= nn//nrow(subset);
   sum= subset[+,];
   mean= mean//sum/n;
   xpx= subset`*subset - sum`*sum/n;
   s= xpx/(n-1);
   s_st= s_st//(n-1)*u_i1`*s*u_i1;
 x_pop= diag(nn);
 free subset;
end;

/*************************************/
/* Generate Comparison Matrices to Compute Sigma_st and use these matrices and the information ob-   */
/* tained using the do-loop to generate the pooled sigma_st matrix */
/*************************************/

if p= 2 then
   do;
     a1= shape({1}, p-1, m);
     pool1= a1*s_st;
     sigma_st= pool1/(n1-g);
   end;
 else if p= 3 then
   do;
     a1= shape({1 0}, p-1, m);
     a2= shape({0 1}, p-1, m);
     pool1= a1*s_st; pool2= a2*s_st;
     sigma_st= (pool1[1,]//pool2[2,])/(n1-g);
   end;
 else if p=4 then
   do;
     a1= shape({1 0 0}, p-1, m);
     a2= shape({0 1 0}, p-1, m);
     a3= shape({0 0 1}, p-1, m);
     pool1= a1*s_st; pool2= a2*s_st; pool3= a3*s_st;
     sigma_st= (pool1[1,]//pool2[2,]//pool3[3,])/(n1-g);
   end;

/*************************************/
/* Complete the necessary computations for the within groups and one-between one-within   */
/* groups repeated measures designs for the multivariate and univariate cases */
/*************************************/

sigma= u_i1`*covmat*u_i1;
m_sigma= sigma_st*(n1-g);
eval1= eigval(sigma_st);
epsilon= (sum(eval1)*sum(eval1))/((p-1)*eval1`*eval1);
theta1= cmatrix1*mean*u_i1;
delta1= theta1`*inv(cmatrix1*inv(x_pop)*cmatrix1`)*theta1;
theta= cmatrix2*mean*u_i1;
delta_st= theta`*inv(cmatrix2*inv(x_pop)*cmatrix2`)*theta;
delta= sqrt(trace(delta_st))/sqrt(trace(sigma_st));
hlt= trace(delta_st*inv(m_sigma));
hlt1= trace(delta1*inv(m_sigma));
s= a><b;
m_within= (hlt1/b)/(1/(n1-p-g+2));
m_inter= ((hlt/s)/(a*b))/(1/(s*(n1-g-b-1)+2));
f_within= trace(delta1)/trace(sigma_st);
f_inter= (trace(delta_st)/(a*b))/(trace(sigma_st)/(b));
rho= min(corrmat);
m_ndf= (p-1)*(g-1);
m_ddf= s*((n-g) - (p-1)-1 + 2);
m_ncp= (m_ddf/s)*hlt;
 if m_ncp >= 50 then m_ncp= 50;
m_fcrit= finv(1-alpha, m_ndf, m_ddf);
m_pwr= 1 - probf(m_fcrit, m_ndf, m_ddf, m_ncp);
gg_ndf= (p-1)*(g-1)*epsilon;
gg_ddf= (p-1)*(n-g)*epsilon;
gg_ncp= b*epsilon*trace(delta_st)/trace(sigma_st);
gg_fcrit= finv(1-alpha, gg_ndf, gg_ddf);
gg_pwr= 1 - probf(gg_fcrit, gg_ndf, gg_ddf, gg_ncp);
```

```
do n2= 12 to 1000 by 3 until (rm_pwr > power);
rm_ndf= (p-1)*(g-1);
rm_ddf= s*(n2-(p-1)-1) + 2;
rm_ncp= ((n2/g)*(delta**2)/2)/(2*(1-rho));
rm_fcrit= finv(1-alpha, rm_ndf, rm_ddf);
rm_pwr= 1 - probf(rm_fcrit, rm_ndf, rm_ddf,
rm_ncp);
end;

lb_eps= 1/(p-1);
do eps1_str= lb_eps to 0.999 by .001 until
(rgg_pwr >= rm_pwr);
 rgg_ndf= (p-1)*(g-1)*eps1_str;
 rgg_ddf= (p-1)*(n2-g)*eps1_str;
 rgg_ncp= rm_ncp*eps1_str;
 rg_fcrit= finv(1-alpha, rgg_ndf, rgg_ddf);
 rgg_pwr= 1 - probf(rg_fcrit, rgg_ndf, rgg_ddf,
rgg_ncp);
end;

do n3= 12 to 1000 by 3 until (pm_pwr > power);
 pm_ndf= (p-1)*(g-1);
 pm_ddf= s*(n3-(p-1)-1) + 2;
 pm_ncp= ((n3/g)*(pdelta**2)/2)/(2*(1-rho));
 pm_fcrit= finv(1-alpha, pm_ndf, pm_ddf);
 pm_pwr= 1-probf(pm_fcrit, pm_ndf, pm_ddf,
pm_ncp);
end;

total_n= n3;
grp_size= total_n/g;
do eps_star= lb_eps to 1.0 by .001 until
(pgg_pwr >= pm_pwr);
 pgg_ndf= (p-1)*(g-1)*eps_star;
 pgg_ddf= (p-1)*(n3-g)*eps_star;
 pgg_ncp= pm_ncp*eps_star;
 pg_fcrit= finv(1-alpha, pgg_ndf, pgg_ddf);
 pgg_pwr= 1 - probf(pg_fcrit, pgg_ndf, pgg_ddf,
pgg_ncp);
 if (pm_pwr > power) & (pgg_pwr >= pm_pwr)
then do;
 end;
end;

/* Generate E_Max and E_Min for Bias Ratio */

  B= P - 1;
  Q= P + 1;
  E1= 1/B;
  E2= (B - 1)/B;
  E3= 1/2*(1 + SQRT(1/B));
  E_MAX= 1/2*(1 + SQRT(B));
  E_MIN= 1/2*(1 - (B - 2)*SQRT(1/B));
  IF E_MIN < .00 THEN E_MIN= .00;

  /* GENERATE THE UNIVARIATE UPPER
BOUND ESTIMATE  */

  U_NDF= (p-1)*(g-1)*EPSILON;
  U_DDF= (p-1)*(n3-g)*EPSILON;
  U_NCP= PM_NCP*e_max;
  U_FCRIT=     FINV(1-ALPHA,     U_NDF,
U_DDF);
  *U_FCRIT= 7.85;
  UU_PWR= 1 - PROBF(U_FCRIT, U_NDF,
U_DDF, U_NCP);

  /* GENERATE THE UNIVARIATE LOWER
BOUND ESTIMATE  */
  U_NDF= (p-1)*(g-1)*EPSILON;
  U_DDF= (n-1)*(n3-g)*EPSILON;
  U_NCP= PM_NCP*e_min;
  U_FCRIT=     FINV(1-ALPHA,     U_NDF,
U_DDF);
  LU_PWR= 1 - PROBF(U_FCRIT, U_NDF,
U_DDF, U_NCP);
  print
'************************************';
  print ' ';
  print '                    Power Analysis Results
';
  print ' ';
  print 'Retrospective: ' Study_n delta rho
gg_pwr m_pwr epsilon eps1_str '                 ';
  print '                    ' f_within f_inter m_within
m_inter '                 ';
  print ' ';
  print '              Prospective: ' total_n grp_size
pdelta epsilon pm_pwr pgg_pwr lu_pwr uu_pwr
'            ';
  print ' ';
  print

'************************************';
%mend powerint;
%powerint(pdelta= .50, power= .80, alpha= .05);
quit;
```