1996

Power Analysis for the Mixed Linear Model

Cheryl Annette Dixon
caddixon@peoplepc.com

Virginia Commonwealth University
School of Medicine

This is to certify that the dissertation prepared by Cheryl Annette Dixon entitled *Power Analysis for the Mixed Linear Model*, has been approved by her committee as satisfactory completion of the dissertation requirement for the degree of Doctor of Philosophy.

[redacted]
R.K. Elswick, Jr., Ph.D., Director of Dissertation

[redacted]
Sunni Choi, Ph.D., School of Medicine

[redacted]
I-li Lu, Ph.D., School of Medicine

[redacted]
Joanne Meyer, Ph.D., School of Medicine

[redacted]
Jean Wyman, Ph.D., School of Nursing

[redacted]
Walter H. Carter, Jr., Ph.D., Department Chairman

[redacted]
Hermes A. Kontos, M.D., Ph.D., Dean, School of Medicine

[redacted]
Jack L. Haar, Ph.D., Dean, School of Graduate Studies

August 5, 1996
Date

Power Analysis for the Mixed Linear Model

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Medical College of Virginia, Virginia Commonwealth University.

by

Cheryl Annette Dixon
B.S., Radford University
Radford, Virginia

Director
Dr. R.K. Elswick, Jr.
Assistant Professor
Department of Biostatistics

Medical College of Virginia
Virginia Commonwealth University
Richmond, Virginia

August, 1996

Acknowledgments

First, I would like to thank my advisor, R.K. Elswick, Jr., for his guidance and encouragement. In particular, I thank him for listening to my constant "I can't do this", and for allowing me to take time out during the writing of my dissertation to spend time with my family. I would also like to thank the rest of my committee members: Sung Choi, I-li Lu, Joanne Meyer, and Jean Wyman, for their time and interest. In addition, I express a special thanks to Vern Chinchilli for his long distance advice and help.

Next, a very special thanks to my mom and dad, Kathy and Floyd, for their continued support and belief in my ability to succeed. Thanks also to my sister, Jennifer, and my future brother-in-law, Scott, for their support and friendship.

Finally, I'd like to thank the friends at MCV that I have made over the years: Kathy, Wendy, Carol, Yvonne, and Shirley. Special thanks to Dianne for making the years at MCV easier to get through, especially the first couple; and to Jenna for making the last couple of years more bearable.

Table of Contents

List of Tables

List of Figures

List of Symbols

α	type I error
$\mathbf{b}$	least squares solution for $\boldsymbol{\beta}$
$\boldsymbol{\beta}$	fixed effects parameters
$\hat{\boldsymbol{\beta}}$	estimate of $\boldsymbol{\beta}$
β	type II error
c	number of columns in a matrix
C	constant
$\mathbf{C}$	g-inverse of coefficient matrix
$\mathbf{C}_{11}$	upper left hand portion of $\mathbf{C}$
$\mathbf{C}_{12}$	upper right hand portion of $\mathbf{C}$
$\mathbf{C}_{21}$	lower left hand portion of $\mathbf{C}$
$\mathbf{C}_{22}$	lower right hand portion of $\mathbf{C}$
δ	noncentrality parameter for t statistics
Δ	small positive number
$\boldsymbol{\varepsilon}$	random error term
η	measure of multivariate association
f	total number of unique matrices in $\mathbf{X}$ and $\mathbf{Z}$
f($\mathbf{y}$,$\mathbf{U}$)	density function
g	constant
g($\mathbf{y}$|$\mathbf{U}$)	density function

$\mathbf{G}$	Var[**U**]
$\ddot{\mathbf{G}}$	essence **G** matrix
$\mathbf{G}^*$	a block of **G**
$\boldsymbol{\gamma}$	vector of unknown parameters
$\boldsymbol{\gamma}^\circ$	an initial estimate of $\boldsymbol{\gamma}$.
h(**U**)	density function
i	subscript
I(*)	identity matrix *×*
j	subscript
k	subscript
K	contrast matrix of fixed effects
ll( )	log likelihood function
L( )	likelihood function
l	number of columns in **L**
L	contrast matrix of fixed and random effects
λ	noncentrality parameter for F statistics
λ^*	primary noncentrality parameter
λ^*_{I}	primary noncentrality parameter for Case I
λ^*_{II}	primary noncentrality parameter for Case II
λ^*_{III}	primary noncentrality parameter for Case III
m	number of unique matrices in Case I essence matrices
μ	mean
$\hat{\mu}$	sample mean
n	number of observations
N	total number of observations
N_e	number of rows in essence design matrix

υ	degrees of freedom
ω	multivariate noncentrality parameter
p	number of fixed effect parameters
Π	power
q	number of random effect parameters
$\mathbf{R}$	Var[$\boldsymbol{\varepsilon}$]
$\ddot{\mathbf{R}}$	essence $\mathbf{R}$ matrix
$\mathbf{R}^{*}$	a block of $\mathbf{R}$
$\mathbf{R}^{*}_{M}$	largest block in $\mathbf{R}$ for Case III
ρ	correlation
S_E	sums of squares error
S_H	sums of squares hypothesis
S^{*}_H	essence sums of squares hypothesis
σ	standard deviation
$\hat{\sigma}$	sample standard deviation
σ^2	variance
$\hat{\sigma}^2$	sample variance
$\boldsymbol{\Sigma}$	var[$\mathbf{y}$]
$\hat{\boldsymbol{\Sigma}}$	estimate of $\boldsymbol{\Sigma}$
$\mathbf{U}$	random effects parameters
w_j	element of $\mathbf{W}$
$\mathbf{W}$	weight matrix
$\mathbf{X}$	fixed effects design matrix
$\ddot{\mathbf{X}}$	fixed effects essence matrix
$[\mathbf{X}^{**}\ \mathbf{Z}]$	overall essence matrix
$\mathbf{X}^{*}$	unique matrix in Case I $\ddot{\mathbf{X}}$

$\otimes$	Kronecker product
$\mathbf{y}$	vector of responses
$\mathbf{Y}$	matrix of responses
$\tilde{\mathbf{Y}}$	transformed responses
$\mathbf{Z}$	random effects design matrix
$\ddot{\mathbf{Z}}$	random effects essence matrix
$\mathbf{Z}^{*}$	unique matrix in Case I $\ddot{\mathbf{Z}}$
$\mathbf{Z}^{\circ}$	common matrix in Case II for $\mathbf{Z}$

Abstract

POWER ANALYSIS FOR THE MIXED LINEAR MODEL

By Cheryl Annette Dixon

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 1996

Major Director: R. K. Elswick, Jr., Ph.D.
Department of Biostatistics

Power analysis is becoming standard in inference based research proposals and is used to support the proposed design and sample size. The choice of an appropriate power analysis depends on the choice of the research question, measurement procedures, design, and analysis plan. The "best" power analysis, however, will have many features of a sound data analysis. First, it addresses the study hypothesis, and second, it yields a credible answer.

Power calculations for standard statistical hypotheses based on normal theory have been defined for t-tests through the univariate and multivariate general linear models. For these statistical methods, the approaches to power calculations have been presented based on the exact or approximate distributions of the test statistics in question. Through the methods proposed by O'Brien and Muller (1993), the noncentrality parameter for the noncentral distribution of the test statistics for the univariate and multivariate general linear models is expressed in terms of its distinct components. This in turn leads to methods for calculating power which are efficient and easy to implement.

As more complex research questions are studied, more involved methods have been proposed to analyze data. One such method includes the mixed linear model. This research extends the approach to power calculation used for the general linear model to the mixed linear model. Power calculations for the mixed linear model will be based on the approximate F statistic for testing the mixed model's fixed effects proposed by Helms (1992). The noncentrality parameter of the approximate noncentral F for the mixed model will be written in terms of its distinct components so that a useful and efficient method for calculating power in the mixed model setting will be achieved. In this research, it has been found that the rewriting of the noncentrality parameter varies depending on study design. Thus, the noncentrality parameter for three specific cases of study design are derived.

Chapter 1
Introduction

1.1 Introduction

The proper planning of an experiment should always include a power analysis. It is becoming a standard addition for inference based research proposals to include power analyses to support the proposed design and sample size (O'Brien and Muller, 1993). Many reviewers of research proposals are now requiring that power analyses be performed before they will recommend funding. The choice of an appropriate power analysis depends on the choice of the research question, measurement procedures, design, and analysis plan. The "best" power analysis, however, will have many features of a sound data analysis. First, it addresses the study hypothesis, and second, it yields a credible answer.

Overestimating or underestimating power can occur when power is computed using approximations of the study design and test statistic rather than the methods for the appropriate design and test. Therefore, the power analysis and statistical analysis must be aligned (Muller et al., 1992). Otherwise, one can commit a type III error. This type of error was termed by Kimball (1957). It is an error which provides the right answer to the wrong question. For example, if an ANOVA is planned and the power is computed for a t-test instead of the power of the ANOVA, then a type III error has been committed.

Performing a power analysis ensures the interaction of statisticians with the researchers. Statisticians who perform a power analysis are more likely to thoroughly familiarized themselves with the proposed design, assess issues regarding data collection

and management, and develop a sound plan for the data analysis. Power calculations can also suggest the need for changes in the proposed study design. Through a power analysis, evaluation of the tradeoffs among type I error rate, type II error rate, choice of variables, choice of analysis, and choice of tests can be made.

Power calculations for standard statistical hypotheses based on normal theory have been defined for t-tests through the univariate and multivariate general linear models. O'Brien and Muller (1993) propose a unified method of power analysis for the t-test through the multivariate hypothesis. Even though power analysis based on many methods is available, O'Brien and Muller present their method because they develop strong parallels between ordinary data analysis and power analysis. Instead of focusing on power of traditional tests, their method allows for the calculation of power for statistical hypotheses that are geared toward more specific research questions. Their approach to power calculation also forces researchers to give specific conjectures or estimates for the relevant parameters to be used in the statistical analysis. These parameters may include population means and standard deviations. Therefore, the results of the power analysis will be improved.

Many of the common test statistics have nonnull distributions that can easily be characterized by either exact or approximate noncentral distributions. The concepts developed by O'Brien and Muller can be applied to all of the cases of the univariate and multivariate general linear models. The basis of their unified approach to power analysis is understanding noncentrality in general linear model testing and knowing how to perform the calculations easily. This leads to expressing the noncentrality parameter in terms of its distinct components.

As more complex research questions are studied, methods of statistical analysis have changed, and consequently, more involved methods have been proposed to analyze data. One such model includes the mixed linear model. The goal of this research is to

extend O'Brien and Muller's work in the univariate and multivariate general linear model to the mixed linear model. For this research, power based on hypothesis testing of the mixed model's fixed effects will be considered. Therefore, the noncentrality parameter based on the approximate F statistic for testing the mixed model's fixed effects, as described by Helms (1991), will be written in terms of its distinct components. Through this, an approach that is useful and efficient for calculating power in the mixed linear model setting will be illustrated.

1.2 Prospectus

Chapter 2 begins with a review of power calculations for the t-test, the univariate general linear model, and the multivariate general linear model. These methods discuss O'Brien and Muller's unified approach and the rewriting of the noncentrality parameter. Following the discussion of power, an example will be presented for each statistical method. In Chapter 3, a review of the mixed linear model will be presented. The chapter begins by briefly discussing the general linear model. Next, mixed linear model methodology including estimation and inference is discussed. The chapter ends by discussing applications of the mixed linear model. Chapter 4 presents the approach for calculating power extended to the mixed linear model setting. The chapter first introduces power for the mixed linear model, in general. The remaining sections of the chapter discuss the rewriting of the noncentrality parameter for three specific cases. These cases arise due to the different study designs that the mixed linear model may be used to analyze. Chapter 5 contains applications of the results from Chapter 4. For each noncentrality case, two examples are presented. (Appendix A contains the programs used to calculate power for each example.) Chapter 6 contains simulation study results concerning the effect of misspecification and underspecification of the mixed linear model's covariance structure. (Appendix B contains the program used for the simulation study.) Chapter 7 focuses on

summary comments and future research ideas concerning power for the mixed linear model.

Chapter 2
Performing a Power Analysis

2.1 What is Power?

Power analyses are now becoming a common requirement for hypothesis-based research proposals. What exactly is power? First, hypothesis testing needs to be understood. A hypothesis is a statement about a population. The goal of a hypothesis test is to determine, based on a sample from the population, which of two hypotheses stated are true. The first hypothesis stated is called the null hypothesis, denoted as H_0. In classical hypothesis testing, the null hypothesis is a statement of "no effect" or "no difference" that one is willing to assume. The second hypothesis is the alternative hypothesis, denoted as H_1. The alternative hypothesis is a statement one hopes or suspects is true instead of H_0 and is considered as the negation of the null hypothesis. A hypothesis test is then a rule that specifies a) for which sample values the decision is made to accept H_0 as true and b) for which sample values H_0 is rejected and H_1 is accepted as true (Casella and Berger, 1990).

When performing a hypothesis test, two decisions can be drawn from the observed data. Each of these decisions can be either correct or incorrect depending on the true situation. Usually, hypothesis tests are evaluated and compared through the probabilities of making errors or incorrect decisions (Casella and Berger, 1990). The following figure describes the decisions that can be drawn in hypothesis testing and their corresponding probabilities.

		Decision	
		Fail to Reject H_0	Reject H_0
True Hypothesis	H_0	Correct Decision $1-\alpha$	Type I error α
	H_1	Type II error β	Correct Decision Power = $1-\beta$

Figure 2.1: Types of Decisions and Their Probabilities in Statistical Hypothesis Testing

The first of the two errors one can commit is a type I error and has probability α. This error is committed when the hypothesis test incorrectly rejects H_0. A type II error is when the hypothesis test incorrectly fails to reject H_0 and has probability β. On the other hand, one can correctly reject H_0. The probability of this is $1-\beta$ and is called power. Therefore, power is the probability of rejecting H_0 (or claiming H_1 is true) given H_1 is in fact true.

The remaining sections of this chapter will review calculating power for various tests of statistical hypotheses. Section 2.2 will review power for the t-test comparing two independent means. Both directional and nondirectional hypotheses will be reviewed. Section 2.3 will discuss power for the univariate general linear model. Both of these sections are based on O'Brien and Muller's (1993) unified discussion of power. Power for the multivariate general linear model will be reviewed in Section 2.4. After power is discussed in each section, an example will be presented using the method discussed.

2.2 Power of t-tests

Common t-tests are probably the most frequently used statistical method. When comparing the population means from two independent groups, the null hypothesis of interest is

$$H_0 : \mu_1 = \mu_2. \tag{2.2.1}$$

The alternative hypothesis can be either directional,

$$H_1 : \mu_1 > \mu_2 \text{ or } H_1 : \mu_1 < \mu_2$$

or nondirectional

$$H_1 : \mu_1 \neq \mu_2.$$

The test statistic for detecting a difference between the sample means is

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}} \tag{2.2.2}$$

where

$\hat{\mu}_i$ is the sample mean for group i,

n_i is the sample size for group i,

and

$\hat{\sigma}$ is the pooled sample standard deviation.

If $N = n_1 + n_2$, then $w_i = n_i/N$ is the proportion of the total sample size in group i. Using these terms, the t statistic given in 2.2.2 can be rewritten as

$$t = (Nw_1w_2)^{1/2}\left(\frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}}\right). \tag{2.2.3}$$

The term $\frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}}$ is known as the effect size.

When H_0 is true, and the observations are independent and follow a normal distribution, then the test statistic, t, has an exact t-distribution with N-2 degrees of freedom. This central t-distribution is denoted as t(N-2). When H_1 is true, t follows a noncentral t-distribution with N-2 degrees of freedom and noncentrality parameter δ, denoted as t(N-2, δ) where

$$\delta = (Nw_1w_2)^{1/2}\left(\frac{\mu_1 - \mu_2}{\sigma}\right). \qquad (2.2.4)$$

Power of the directional test, using $H_1 : \mu_1 > \mu_2$, is

$$\Pi = P[t(N-2, \delta) \geq t_\alpha] \qquad (2.2.5)$$

where t_α is the upper-tail critical value satisfying

$$\alpha = P[t(N-2) \geq t_\alpha]$$

and α is the type I error rate. In other words, α is the probability that the random t-variate with N-2 degrees of freedom will exceed the critical value, t_α, given δ=0, and power is the probability that t exceeds t_α but given the noncentrality parameter in 2.2.4 which is a value greater than zero. If $H_1 : \mu_1 < \mu_2$ is being tested, then $-t_\alpha$ would be used in place of t_α in the above statements.

Nondirectional tests, i.e., $H_1 : \mu_1 \neq \mu_2$, use $t_{\alpha/2}$ and $-t_{\alpha/2}$. It is, however, more straightforward to use the fact that $F_\alpha = t^2_{\alpha/2}$, so that the split rejection region of the central t-

distribution is unified into the upper tail of the central F-distribution (O'Brien and Muller, 1993). Therefore, H_0 can be tested by using the test statistic $F=t^2$. Under H_0, F follows a central F-distribution with a single numerator degree of freedom and N-2 denominator degrees of freedom. This central F is denoted as F(1, N-2). Under H_1, F follows a noncentral F-distribution with a single numerator degree of freedom, N-2 denominator degrees of freedom, and noncentrality parameter, λ. This noncentral F is denoted as $F(1, N-2, \lambda)$ where $\lambda=\delta^2$. Power for the nondirectional (two-tailed) t-test is then

$$\Pi=P[F(1, N-2, \lambda) \geq F_\alpha] \qquad (2.2.6)$$

where F_α is the upper tail critical value satisfying

$$\alpha=P[F(1, N-2) \geq F_\alpha]$$

and α is the type I error rate. In other words, α is the probability that the random F-variate with 1 numerator degree of freedom and N-2 denominator degrees of freedom will exceed the critical value F_α given $\lambda=0$, and power is the probability that F exceeds F_α but given the particular $\lambda>0$.

2.2.1 Illustration of t-test Power

To illustrate these methods, suppose a surgeon is planning a study in which he wants to compare the mean number of days a patient remains in the hospital following one of two surgical procedures. Even though he has no pilot data, he is willing to assume that the first procedure has a mean of 9 days and he believes the effect of the second procedure is a 33% reduction or a mean of 6 days. He is also willing to assume a common within

group standard deviation as small as 4 days and as large as 5.5 days. He is planning a balanced design, i.e., equal sample sizes or $w_1=w_2=0.5$. For testing the alternative hypothesis, $H_1 : \mu_1 > \mu_2$, he is interested in powers for the two standard deviations of 4 and 5.5 at $\alpha=.05$. Since he performs approximately 40 such surgical procedures a year, he is hoping he will have sufficient power by studying a total of N=40 subjects, and consequently, finish the study in one year.

The following plot shows the power curves for the two standard deviations of 4 and 5.5.

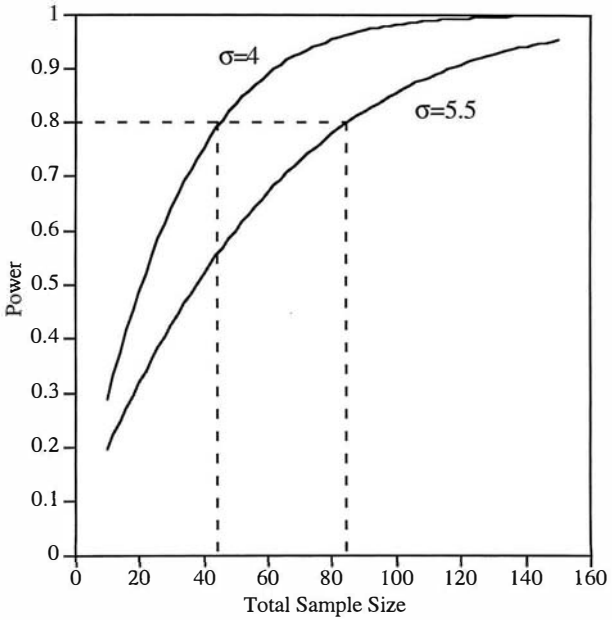

Figure 2.2: Plot of Power Curves for t-test Example

From the plot, one can see that the surgeon will need a total of 46 subjects or 23 subjects per group to achieve a power of 0.80 with a standard deviation of 4. He will need a total of 86 subjects or 43 subjects per group with a standard deviation of 5.5 to achieve the same power of 0.80. The surgeon is not sure if he can get the total of 46 subjects in one year. He would still like to use only 40 subjects, so he has decided to rethink his assumptions and return to the statistician at a later date.

2.3 Power for the Univariate General Linear Model

Consider the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$\mathbf{y}$ is the N×1 vector of responses,

$\mathbf{X}$ is the N×p known full-rank fixed effects design matrix,

$\boldsymbol{\beta}$ is the p×1 unknown vector of fixed effects parameters,

and

$\boldsymbol{\varepsilon}$ is the N×1 vector of random errors.

For tests on $\boldsymbol{\beta}$, it will be assumed that the elements of $\boldsymbol{\varepsilon}$ are independent $N(0, \sigma^2)$ random variables. The usual estimates for $\boldsymbol{\beta}$ and σ^2 are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

and

$$\hat{\sigma}^2 = \frac{\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{N - p}$$

Assume that testing the general linear hypothesis,

$$H_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0} \qquad (2.3.1)$$
$$H_1 : \mathbf{L}'\boldsymbol{\beta} \neq \mathbf{0},$$

is of interest. In 2.3.1, L is a p×l ,with rank(**L**)=l ≤p, matrix of contrasts. These contrasts are the combinations of the population means that are of interest. The test statistic for testing H_0 is

$$F = \frac{\left(\mathbf{L}'\hat{\boldsymbol{\beta}}\right)'\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}\left(\mathbf{L}'\hat{\boldsymbol{\beta}}\right)}{l \cdot \hat{\sigma}^2}. \qquad (2.3.2)$$

The term in the numerator of the F statistic is known as the sum of squares for the hypothesis. Under H_0, the F statistic has an exact F-distribution with l numerator degrees of freedom and N-p denominator degrees of freedom, denoted as F(l, N-p). Under H_1, F follows a noncentral F-distribution with l numerator degrees of freedom, N-p denominator degrees of freedom, and noncentrality parameter, λ. The noncentral F is denoted as F(l, N-p, λ) where

$$\lambda = \frac{(\mathbf{L}'\boldsymbol{\beta})'\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}(\mathbf{L}'\boldsymbol{\beta})}{\sigma^2}. \qquad (2.3.3)$$

Power for the above hypothesis is then

$$\Pi = P[F(l, \text{N-p}, \lambda) \geq F_\alpha] \qquad (2.3.4)$$

where F_α is the upper tail critical value satisfying

$$\alpha = P[F(l, N-p) \geq F_{\alpha}].$$

The noncentrality parameter, λ as given in 2.3.2, can be expressed in terms of its distinct components as discussed in O'Brien and Muller (1993). Let $\ddot{\mathbf{X}}$ be the n×p essence model matrix formed by assembling the n unique rows of $\mathbf{X}$. That is, $\ddot{\mathbf{X}}$ contains the unique design points for the study of interest. Let $\mathbf{W}$ be the n×n diagonal matrix with elements w_j where $0<w_j<1$ and $\sum w_j=1$. The elements, w_j, are the proportion of the total sample size associated with the j^{th} row of $\ddot{\mathbf{X}}$. In other words, $\mathbf{X}$ has Nw_j rows identical to the j^{th} row of $\ddot{\mathbf{X}}$. Therefore, N$\mathbf{W}$ holds the n sample sizes. Since

$$\mathbf{X}'\mathbf{X} = N\left(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}}\right),$$

the noncentrality parameter can now be written as

$$\begin{aligned}
\lambda &= \frac{(\mathbf{L}'\boldsymbol{\beta})'\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}(\mathbf{L}'\boldsymbol{\beta})}{\sigma^2} \\
&= \frac{(\mathbf{L}'\boldsymbol{\beta})'\left[\mathbf{L}'\left(N\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}}\right)^{-1}\mathbf{L}\right]^{-1}(\mathbf{L}'\boldsymbol{\beta})}{\sigma^2} \\
&= \frac{N(\mathbf{L}'\boldsymbol{\beta})'\left[\mathbf{L}'\left(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}}\right)^{-1}\mathbf{L}\right]^{-1}(\mathbf{L}'\boldsymbol{\beta})}{\sigma^2} \\
&= N\lambda^*
\end{aligned}$$

where λ^* is called the primary noncentrality parameter. It is seen that λ^* is not based on N, the total sample size. It is based solely on the design points to be used ($\ddot{\mathbf{X}}$), the sample weightings of those points ($\mathbf{W}$), and the conjectured values for $\boldsymbol{\beta}$ and σ^2. Thus, when calculating power for various total sample sizes, the primary noncentrality only needs to be

calculated once. Therefore, instead of calculating λ each time the total sample size differs, using λ^* is a computationally more efficient way of calculating power.

2.3.1 Illustration of Univariate Power

By way of an example, suppose a pharmaceutical company will be sponsoring a two center clinical trial. The goal of the experiment is to test two levels of the treatment, an antihypertensive drug (10 mg and 20 mg) versus a placebo. The response of interest is diastolic blood pressure from moderately hypertensive patients. Using data from a previous study, the clinician expects to see, from the first center, post treatment means of 98, 88, and 82 from placebo, 10 mg, and 20 mg groups, respectively. She also expects that the means for the second center will be 10% lower. Therefore, the means assumed for the second center will be 88.2, 79.0, and 73.8. Using the mean square error from the previous study, an estimate of the variance is assumed to be σ^2=142.3. A cell means model of the form

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \quad i=1,2,3 \quad j=1,2 \quad k=1,2,...,n$$

will be fit. In this model,

y_{ijk} is the response of the k^{th} subject from the i^{th} treatment level from the j^{th} center,

μ_{ij} is the mean response from the i^{th} treatment level at the j^{th} center,

and

ε_{ijk} is the random error term for the k^{th} subject from the i^{th} treatment level from the j^{th} center.

Therefore, in general linear model form, the essence matrix, $\ddot{\mathbf{X}}$, has the form

$$\ddot{\mathbf{X}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} = [98 \quad 88 \quad 82 \quad 88.2 \quad 79.2 \quad 73.8]'.$$

In $\ddot{\mathbf{X}}$ and $\boldsymbol{\beta}$, the first three columns correspond to the first center placebo, 10 mg, and 20 mg dose levels, respectively, and the last three columns correspond to the second center placebo, 10 mg, and 20 mg dose levels. The clinician wants equal weightings for each treatment center combination, therefore,

$$\mathbf{W} = \begin{bmatrix} 1/6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/6 \end{bmatrix}.$$

Since the pharmaceutical company is interested in testing treatment effect, the contrast of interest is

$$\mathbf{L}' = \begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 \end{bmatrix}.$$

The first row of the contrast compares the placebo effect to the effect of the 10 mg dose level from each center and the second row of the contrast compares the placebo effect to the

effect of the 20 mg dose level from each center. Using $\ddot{\mathbf{X}}$, $\boldsymbol{\beta}$, $\mathbf{W}$, σ^2, and $\mathbf{L}$ as defined above, the primary noncentrality is $\lambda^*=0.2762$.

The clinician is hoping she can perform the experiment with 10 or fewer patients per cell and still achieve an adequate power. Upon further consideration, she also thinks that the variance may be inflated by 25% or $1.25\times142.3=177.8$. The primary noncentrality for this variance is $\lambda^*=0.2211$. Notice that with the larger variance, the value of the primary noncentrality parameter decreases. The power curves for both variances at $\alpha=0.05$ are shown in the following plot.

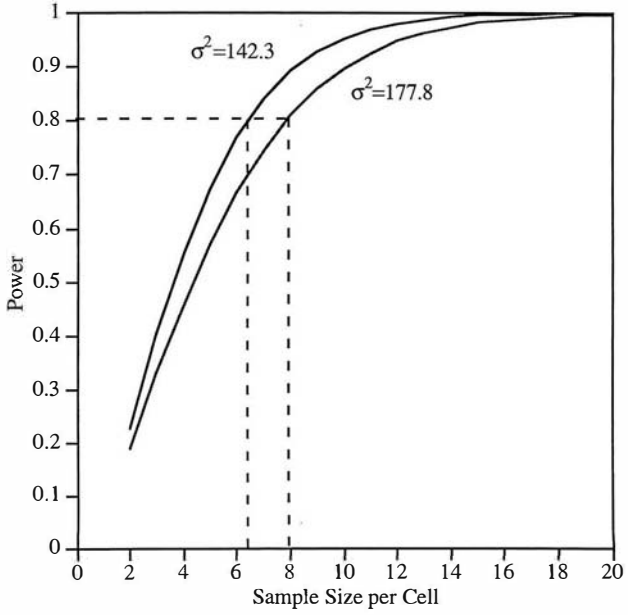

Figure 2.3: Plot of Power Curves for Univariate GLM Example

From the plot, one can see that the clinician, for 80% power, will need 7 subjects per center per treatment or a total of 42 subjects when a variance of 142.3 is assumed. If the 25% inflation of the variance is assumed, she will need 8 subjects per center per treatment or a total of 48 subjects. If she is willing to use the maximum of 10 patients per center per treatment, she will have power closer to 90% assuming the larger variance.

2.4 Power for the Multivariate Linear Model

The standard multivariate linear model has the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$\mathbf{Y} = \left[\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_p\right]$ is the N×p matrix of the responses for the p dependent variables and $\mathbf{y}_i$ is a N×1 vector of responses for the i^{th} dependent variable,

$\mathbf{X}$ is the N×r known full-rank fixed effects design matrix, as in the univariate case,

$\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2 \quad \cdots \quad \boldsymbol{\beta}_p\right]$ is the r×p matrix of unknown fixed effects parameters and each $\boldsymbol{\beta}_i$ is a r×1 vector,

and

$\boldsymbol{\varepsilon} = \left[\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2 \quad \cdots \quad \boldsymbol{\varepsilon}_p\right]$ is the N×p matrix of random errors and each $\boldsymbol{\varepsilon}_i$ is a N×1 vector.

The rows of $\boldsymbol{\varepsilon}$ are assumed to be independent p variate normal random vectors with mean $\mathbf{0}_{p\times 1}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a p×p positive definite matrix. Under the conditions specified above, the usual estimates are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and

$$\hat{\boldsymbol{\Sigma}} = \frac{\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{\mathrm{N} - \mathrm{r}}$$

The multivariate general linear hypothesis is

$$\mathrm{H}_0: \mathbf{C}\boldsymbol{\beta}\mathbf{U} = \mathbf{0}_{c \times u} \tag{2.4.1}$$

where $\mathbf{C}$ is a known, full rank, c×r matrix with c≤r and $\mathbf{U}$ is a known, full rank, p×u matrix with u≤p. Consequently, the degrees of freedom for H_0 are equal to cu.

The matrix $\mathbf{C}$ is used to compare the r design effects, i.e., it controls contrasts on the rows of $\boldsymbol{\beta}$. Each row of $\mathbf{C}$ can be referred to as a between-subject contrast. The matrix $\mathbf{U}$ is used to compare the p responses, i.e., it controls contrasts on the columns of $\boldsymbol{\beta}$. Each row of $\mathbf{U}$ can be referred to as a within-subject contrast. The matrix $\mathbf{U}$ also corresponds to a transformation of the responses. The transformed responses can be written as $\tilde{\mathbf{Y}} = \mathbf{YU}$ and yield the following model

$$\mathbf{YU} = \mathbf{X}\boldsymbol{\beta}\mathbf{U} + \boldsymbol{\varepsilon}\mathbf{U}$$
$$\tilde{\mathbf{Y}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}$$

The rows of $\tilde{\boldsymbol{\varepsilon}}$ are independent u-variate normal random vectors with mean $\mathbf{0}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}} = \mathbf{U}'\boldsymbol{\Sigma}\mathbf{U}$. The null hypothesis of interest then becomes

$$\mathrm{H}_0: \mathbf{C}\tilde{\boldsymbol{\beta}} = \mathbf{0}.$$

If u = 1, then the multivariate general linear model on $\mathbf{Y}$ becomes a univariate general linear model on $\tilde{\mathbf{Y}} = \mathbf{YU}$ and the methods of Section 2.3 apply directly.

When u > 1, the sums of squares for the hypothesis from the univariate general linear model generalizes to

$$\begin{aligned} S_H &= \left(\mathbf{C}\hat{\boldsymbol{\beta}}\mathbf{U}\right)' \left[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}}\mathbf{U}\right) \\ &= N\left(\mathbf{C}\hat{\boldsymbol{\beta}}\mathbf{U}\right)' \left[\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}\right]^{-1} \left(\mathbf{C}\hat{\boldsymbol{\beta}}\mathbf{U}\right) \\ &= N \cdot S_H^*. \end{aligned} \tag{2.4.2}$$

where $\ddot{\mathbf{X}}$ and $\mathbf{W}$ are as defined in Section 2.3, and S_H^* is the essence sums of squares and cross-products matrix for the hypothesis. Thus, S_H is the overall u×u sums of squares and cross-products matrix for the hypothesis.

The variance term, $\hat{\sigma}^2$ in the univariate general linear model, generalizes to $\mathbf{U}'\hat{\boldsymbol{\Sigma}}\mathbf{U} = S_E/(N-r)$, where

$$S_E = \mathbf{U}'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)\mathbf{U}. \tag{2.4.3}$$

S_E follows a central Wishart distribution with N-r degrees of freedom. Under the null hypothesis, S_H follows a central Wishart with c degrees of freedom.

The usual multivariate test statistics can be defined as functions of the eigenvalues of $S_H S_E^{-1}$, of which at most s=min(c, u) are positive. It is also common to express the statistics in terms of the s positive eigenvalues of $S_H(S_H + S_E)^{-1}$, which are the generalized, squared canonical correlations. By expressing the multivariate test statistics as functions of the canonical correlations, multivariate power analysis can be explained in terms of the corresponding univariate results (Muller, et al., 1992). In the univariate case, u=1 and the test statistic considered can be written in terms of the squared multiple correlation, $\hat{\rho}^2$, as

$$F = \frac{\hat{\rho}^2/(\text{df model})}{(1-\hat{\rho}^2)/(\text{df error})}. \qquad (2.4.4)$$

Thus, the multivariate test statistics can similarly be expressed as functions of the measure of multivariate association, $\hat{\eta}$, and are

$$F = \frac{\hat{\eta}/(cu)}{(1-\hat{\eta})/\upsilon} \qquad (2.4.5)$$

where υ is the denominator degrees of freedom. The measure of multivariate association is a function of the multivariate test statistic of interest which is a function of the generalized canonical correlations.

The four most common multivariate test statistics are Roy's largest root (RLR), Wilk's likelihood ratio statistic (WLR), Pillai-Bartlett trace (PBT), and Hotelling-Lawly trace (HLT). The last three statistics use all s of the eigenvalues of $S_H S_E^{-1}$ but RLR uses only the largest eigenvalue. All four statistics provide a size α test, but they are equivalent only if s=min(c, u)=1. Even though the statistics are easy to compute, the associated p-values are not. There are no general exact formulas for the distribution functions under the null hypothesis. However, WLR, PBT, and HLT can all be converted into approximate F statistics.

The Wilk's likelihood ratio statistic is the determinant of $S_E(S_H + S_E)^{-1}$ or $\text{WLR} = \left|S_E(S_H + S_E)^{-1}\right|$. Rao's transformation converts this to an F statistic with cu numerator degrees of freedom and υ_{WLR} denominator degrees of freedom,

$$F_{WLR} = \upsilon_{WLR} \cdot \left(\frac{1-\text{WLR}^{1/t}}{\text{WLR}^{1/t}}\right)\Big/ cu \qquad (2.4.6)$$

where

$$g = \begin{cases} 1 & cu \le 3 \\ \left[\left(c^2u^2 - 4\right)/\left(c^2 + u^2 - 5\right)\right]^{1/2} & cu \ge 4 \end{cases}$$

and

$$\upsilon_{WLR} = g\left[N - r - (u - c + 1)/2\right] - (cu - 2)/2.$$

The multivariate measure of association based on WLR is

$$\hat{\eta}_{WLR} = 1 - WLR^{1/g}. \tag{2.4.7}$$

Substituting $\hat{\eta}_{WLR}$ and υ_{WLR} into equation 2.4.5 leads to equation 2.4.6.

The Pillai-Bartlett trace statistic is based on the trace of $S_H\left(S_H + S_E\right)^{-1}$ or $PBT = tr\left[S_H\left(S_H + S_E\right)^{-1}\right]$. The transformation of this into an F statistic with cu numerator degrees of freedom and υ_{PBT} denominator degrees of freedom is

$$F_{PBT} = \upsilon_{PBT} \cdot \left(\frac{PBT}{s - PBT}\right) \Big/ cu \tag{2.4.8}$$

where $\upsilon_{PBT} = s[N - r - u - s]$. For PBT, the multivariate measure of association is

$$\hat{\eta}_{PBT} = \frac{PBT}{s}. \tag{2.4.9}$$

Upon substituting $\hat{\eta}_{PBT}$ and υ_{PBT} in equation 2.4.5 and simplifying, leads to equation 2.4.8.

The Hotelling-Lawly trace statistic is based on the trace of $S_H S_E^{-1}$ or $HLT = tr\left[S_H S_E^{-1}\right]$. The transformation of this into an F statistic with cu numerator degrees of freedom and υ_{HLT} denominator degrees of freedom is

$$F_{HLT} = \upsilon_{HLT} \cdot \left(\frac{HLT}{s}\right) \Big/ cu \qquad (2.4.10)$$

where $\upsilon_{HLT} = s[N - r - u - 1] + 2$. The multivariate measure of association based on HLT is

$$\hat{\eta}_{HLT} = \frac{HLT/s}{1 + (HLT/s)}. \qquad (2.4.11)$$

Substituting $\hat{\eta}_{HLT}$ and υ_{HLT} into equation 2.4.5, it is easy to show that

$$F_{HLT} = \frac{\hat{\eta}_{HLT}/(c \cdot u)}{(1 - \hat{\eta}_{HLT})/\upsilon_{HLT}}.$$

Power calculations in testing the multivariate general linear hypothesis have presented difficult problems. These problems arise from the fact that exact general formulas are not available for any of the four common multivariate test statistics in the null case, and therefore, closed form expressions are not available for distributions of the multivariate test statistics under the alternative hypothesis. However, practical asymptotic approximations based on mixtures of noncentral chi-squares have been available since the early 1970's for WLR, PBT, and HLT. Sugiura and Fujikoski (1969) provided a general approximation for noncentral probabilities for WLR. Lee (1971) presented general approximations for WLR, PBT, and HLT. These approximations involve asymptotic approximation of the characteristic function, followed by inversion, which yields a mixture of noncentral chi-squares. Approximately ten chi-square terms with complicated coefficients are needed to define the approximation. Power calculations are then based on

the trace of the noncentrality matrix and traces of its higher powers such as the square of the noncentrality matrix.

The chi-square approximation methods are very complex. Using a generalization of computing power for the univariate general linear hypothesis, Muller and Peterson (1984) suggest noncentral F approximations considering the expressions for the above central F approximations (2.4.6, 2.4.8, 2.4.10). Their F approximations are simpler to implement and simpler to understand. Muller and Peterson suggest that under H_1, the above F statistics follow an approximate noncentral F with cu numerator degrees of freedom, υ_i denominator degrees of freedom, and noncentrality parameter ω_i. For these statistics, υ_i is the corresponding denominator degrees of freedom for WLR, PBT, and HLT and $\omega_i = cu \cdot F_A(i)$. $F_A(i)$ is the F that would be observed if one obtained $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}$, i.e., if one obtained the true values, for the corresponding test statistic WLR, PBT, and HLT. Thus, approximate power for a given test statistic is

$$\Pi = P[F(cu, \upsilon_i, \omega_i) \geq F_\alpha] \qquad (2.4.12)$$

where F_α is the upper tail critical value satisfying

$$\alpha = P[F(cu, \upsilon_i) \geq F_\alpha].$$

The F approximation method of power approximation yields sufficiently accurate results. Muller and Peterson found that the chi-square approximations for WLR, PBT, and HLT lead to nearly three digits of accuracy in computing power. Their F approximations for WLR, PBT, and HLT provide nearly 2 digits of accuracy in computing power which they deem sufficient for almost any practical situation. A general, practical power approximation for Roy's largest root does not exist because even its null distribution is

difficult to characterize as an approximate F. With the F approximation method, a primary noncentrality parameter cannot be defined. O'Brien and Muller (1993), however, mention that O'Brien and Shieh (under review) have proposed a modification to this method that defines a primary noncentrality and may give more accurate results.

2.4.1 Illustration of Multivariate Power

To illustrate the methods presented in this section, suppose a psychologist is interested in comparing the psychological distress between patients who accept enrollment into a drug treatment program and those patients who reject enrollment into the program. It has been seen that those who enroll suffer greater distress and want the help to better themselves. In this study, patients will be administered a written test that will measure four psychological symptoms. Therefore, each subject will have four dependent responses. These responses include the scores for each of the following factors: depression, anxiety, hostility, and phobia. A profile analysis, which is a one-way MANOVA in which the profiles for the two groups will be compared, is planned. In a profile analysis, there are three hypotheses of interest. The first is a test of parallelism. Parallelism corresponds to the lack of a response×group interaction. The other two hypotheses are coincidence of profiles (equivalence of groups) and constancy of profiles (equivalence of responses). Due to budget restraints at this time, the psychologist would like to get by with a total sample size of no more than 90 subjects.

Using data from a current study being done by the Center for Perinatal Addiction, Division of Substance Abuse Medicine from the Medical College of Virginia, the psychologist gets estimates for the mean scores for each of the psychological factors. Subjects who enroll into the drug treatment program have mean scores of 63, 57, 55, and 57 for depression, anxiety, hostility, and phobia, respectively. The subjects who reject

enrollment have mean scores of 58, 50, 55, and 48 for depression, anxiety, hostility, and phobia, respectively. Thus, $\boldsymbol{\beta}$ has the form

$$\boldsymbol{\beta} = \begin{array}{c} \begin{matrix} \text{Dep} & \text{Anx} & \text{Hos} & \text{Phob} \end{matrix} \\ \begin{bmatrix} 63 & 57 & 55 & 57 \\ 58 & 50 & 55 & 48 \end{bmatrix} \end{array} \begin{matrix} \\ \text{Acceptors} \\ \text{Rejectors} \end{matrix}$$

Using the data from the above study, an estimate of the variance-covariance is assumed to be

$$\boldsymbol{\Sigma} = \begin{array}{c} \begin{matrix} \text{Dep} & \text{Anx} & \text{Hos} & \text{Phob} \end{matrix} \\ \begin{bmatrix} 78 & 77 & 56 & 53 \\ 77 & 152 & 71 & 87 \\ 56 & 71 & 143 & 40 \\ 53 & 87 & 40 & 140 \end{bmatrix} \end{array} \begin{matrix} \text{Dep} \\ \text{Anx} \\ \text{Hos} \\ \text{Phob.} \end{matrix}$$

The main hypotheses of interest for the psychologist's research question are the hypotheses of parallelism and coincidence of the profiles. The hypothesis of constancy of equivalence of the responses is not of interest. The null hypothesis of $H_0 : \mathbf{C}\boldsymbol{\beta}\mathbf{U} = \mathbf{0}$ for both tests, parallelism and coincidence, use the following $\mathbf{C}$ matrix

$$\mathbf{C} = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

For the test of parallelism or the response×group interaction, the $\mathbf{U}$ matrix used is

$$\mathbf{U}_P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

The test of coincidence or group effect uses the **U** matrix

$$\mathbf{U}_C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since the number of rows of **C** equals 1 and therefore, s=min(c, u)=min(1,(3 or 4))=1, all of the multivariate tests are equivalent. Therefore, the powers will be calculated with the Wilk's Likelihood Ratio statistic only.

The psychologist hopes to have twice as many subjects enroll into the drug treatment program than reject enrollment. Figure 2.4 and Figure 2.5 contain the plots of the power curves for each hypothesis. The first plot assumes an equal sample size per group and the second plot assumes sample sizes with the 2:1 ratio.

From the plot in Figure 2.4, a total sample size of 90 subjects or 45 subjects per group yield powers of 0.72 and 0.91 for the test of the response×group interaction and group effect, respectively. After much consideration, the psychologist is sure she will be getting twice as many subjects enrolling into the treatment program rather than rejecting the treatment program. Therefore, for a total sample size of 90 subjects or 60 subjects enrolling and 30 subjects rejecting, powers are only 0.66 and 0.87 for parallelism and coincidence, respectively as seen in Figure 2.5.

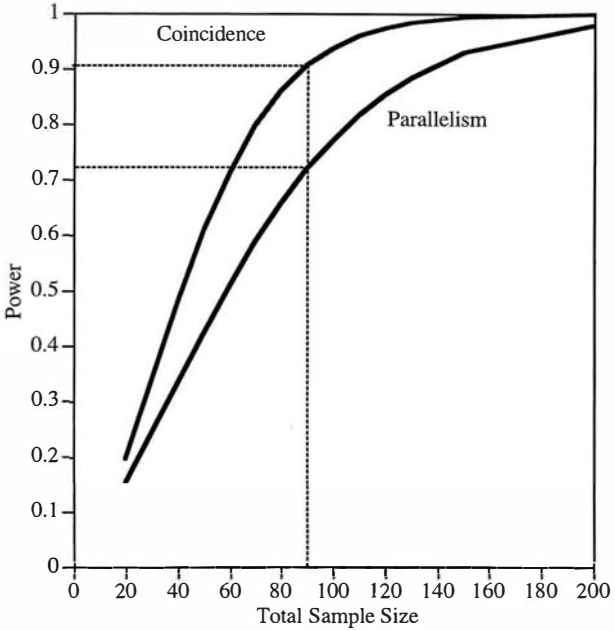

Figure 2.4: Plot of Power Curves for Multivariate GLM Example Assuming Equal Sample Size per Group

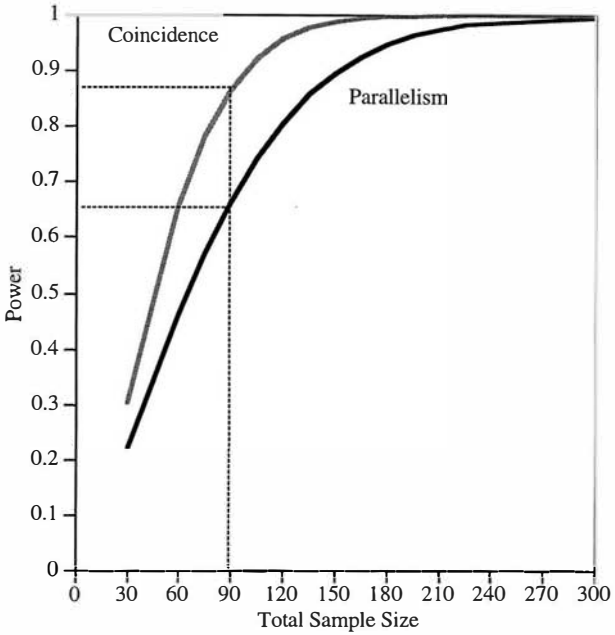

Figure 2.5: Plot of Power Curves for Multivariate GLM Example Assuming Sample Size Ratio of 2 to 1

The psychologist, however, is bothered by the low powers for the response×group interaction test. In order to have a power of at least 0.80 for this test, she will need to increase her total sample size to 110 subjects or 55 subjects per group for equal sample size or increase her total sample size to 120 subjects or 80 subjects enrolled and 40 subjects rejecting. At these total sample sizes, powers for the group effect increase greatly. The psychologist is now going to see if she can find additional funding so that she may sample the total of 120 subjects since she is certain that she will sample twice as many subjects

enrolling into the program than rejecting enrollment. She does realize that if she would be able to sample an equal number of subjects from each group, she would have larger powers. For example, if 120 total subjects are sampled, powers are 0.85 for parallelism and 0.97 for coincidence when equal sample sizes are taken but the powers are only 0.81 for parallelism and 0.95 for coincidence when samples are taken in a 2 to 1 ratio.

Chapter 3
The Mixed Linear Model

3.1 The General Linear Model: A Brief Review

A statistical model is a mathematical description of the mechanism that generates a set of data (Wolfinger, 1992). The most common statistical model is the general linear model (GLM) which has the form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (3.1.1)$$

where

$\mathbf{y}$ is a N×1 vector of the observed responses,

$\mathbf{X}$ is a N×p known design matrix of the fixed effects with rank($\mathbf{X}$)=p,

and without loss of generality, $\mathbf{X}$ is assumed to be full rank,

$\boldsymbol{\beta}$ is a p×1 unknown vector of the fixed effect parameters,

and

$\boldsymbol{\varepsilon}$ is a n×1 unknown vector of random errors.

It is assumed that the components of $\boldsymbol{\varepsilon}$ i.e., the elements ε_i, are independent and normally distributed with mean 0 and common variance σ^2.

The general linear model allows for the modelling of the mean of $\mathbf{y}$ by using the fixed effects, $\boldsymbol{\beta}$, which are estimated. One method used to estimate $\boldsymbol{\beta}$ is the method of least squares. This method finds the value of $\boldsymbol{\beta}$ that minimizes

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

which is the sum of squares of the residuals. This method requires only that the vector $\boldsymbol{\varepsilon}$ of random errors has mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}$. To minimize this expression as a function of $\boldsymbol{\beta}$, differentiate the above expression with respect to $\boldsymbol{\beta}$, set the derivative equal to zero, and the resulting equation, the normal equation,

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{3.1.2}$$

is then solved for $\hat{\boldsymbol{\beta}}$. The least squares estimate is then

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

This estimate is the best linear unbiased estimator (BLUE) for $\boldsymbol{\beta}$. Also, $\hat{\boldsymbol{\beta}}$ can be shown to be the maximum likelihood estimate of $\boldsymbol{\beta}$. The maximum likelihood estimate for $\boldsymbol{\beta}$ is found by maximizing the likelihood function which is defined to be the joint density of the random errors. In this case, it is further assumed that the random errors are independently normally distributed. Thus, the joint density is the product of the marginals and the likelihood function is

$$\begin{aligned} L(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N; \sigma^2) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\tfrac{1}{2}\left(\tfrac{\varepsilon}{\sigma}\right)^2\right\} \\ &= \frac{1}{(2\pi)^{N/2}(\sigma^2)^{N/2}} \exp\left\{\left(-\tfrac{1}{2\sigma^2}\right)\sum_{i=1}^{N}\varepsilon_i^2\right\}. \end{aligned}$$

Since $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $\sum_{i=1}^{N} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$. Substituting this quantity into the previous equation, the likelihood function becomes

$$L(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N; \sigma^2) = \frac{1}{(2\pi)^{N/2}(\sigma^2)^{N/2}} \exp\left\{\left(-\tfrac{1}{2\sigma^2}\right)\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]\right\}.$$

Taking the natural logarithm of each side and simplifying leads to

$$\ln\ L(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N; \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]. \quad (3.1.3)$$

To maximize equation 3.1.3 with respect to $\boldsymbol{\beta}$, ln(L) is differentiated with respect to $\boldsymbol{\beta}$, set equal to zero, and solved for $\boldsymbol{\beta}$. The only term that contains $\boldsymbol{\beta}$ is the third term of the above expression and this is equivalent to the term that was used in the least squares setting. Therefore, given the assumptions on $\boldsymbol{\varepsilon}$, $\hat{\boldsymbol{\beta}}$ is found to be the minimum variance unbiased estimator (MVUE) of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$ is normally distributed with mean $\boldsymbol{\beta}$ and variance $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Since σ^2 is usually unknown, it can be estimated using maximum likelihood techniques. By differentiating equation 3.1.3 with respect to σ^2, setting the result equal to zero, and solving yields the maximum likelihood estimate

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{N}.$$

Adjusted for bias, the maximum likelihood estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)}{\mathrm{N} - \mathrm{rank}(\mathbf{X})}$$

and is preferred over the biased estimate. This term is known as the residual mean square error.

Inferences can be made on the general linear model by forming linear combinations of $\boldsymbol{\beta}$ that are of interest and then using the sampling distribution to form hypothesis tests. Consider estimable linear combinations of the form $\mathbf{L}'\boldsymbol{\beta}$ where $\mathbf{L}$ is a known, full rank, estimable coefficient matrix. The best linear unbiased estimator under least squares for $\mathbf{L}'\boldsymbol{\beta}$ is $\mathbf{L}'\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is the least squares estimator for $\boldsymbol{\beta}$. $\mathbf{L}'\boldsymbol{\beta}$ has an estimated variance of $\hat{\sigma}^2\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}$ since the estimate of the variance of $\hat{\boldsymbol{\beta}}$ is $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$.

If $\mathbf{L}$ is of rank 1, the t-statistic used for testing

$$\mathrm{H}_0 : \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$$

is of the form:

$$\mathrm{t} = \frac{\mathbf{L}'\hat{\boldsymbol{\beta}}}{\sqrt{\hat{\sigma}^2\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}}}.$$

Under the assumed normality of $\boldsymbol{\varepsilon}$, t has an exact t-distribution with υ degrees of freedom where υ=N-rank($\mathbf{X}$).

If $\mathbf{L}$ has rank greater than 1, then a F-statistic is used for testing the above hypothesis and has the form:

$$\mathrm{F} = \frac{\hat{\boldsymbol{\beta}}'\mathbf{L}\left[\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{-1}\mathbf{L}'\hat{\boldsymbol{\beta}}}{\hat{\sigma}^2\mathrm{rank}(\mathbf{L})}.$$

This F-statistic has an exact F-distribution with numerator degrees of freedom rank($\mathbf{L}$) and υ denominator degrees of freedom when $\boldsymbol{\varepsilon}$ is assumed to be normally distributed. For a complete discussion of the general linear model, see Searle (1971).

3.2 Mixed Linear Model Methodology

The mixed linear model is an extension of the general linear model. Written in Henderson's (1984) notation, the mixed model is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon} \qquad (3.2.1)$$

where

$\mathbf{y}$ is a N×1 vector of measured responses,

$\mathbf{X}$ is a N×p known design matrix for the fixed effects with rank $\mathbf{X}$ = p where p ≤ N,

$\boldsymbol{\beta}$ is a p×1 vector of the unknown fixed effect parameters,

$\mathbf{Z}$ is a N×q known design matrix for the random effects,

$\mathbf{U}$ is a q×1 vector of unknown random effects,

and

$\boldsymbol{\varepsilon}$ is a N×1 vector of random errors.

The assumptions made on $\mathbf{U}$ are

$$E[\mathbf{U}] = \mathbf{0}_{q\times 1} \text{ and } \mathrm{Var}[\mathbf{U}] = \mathbf{G} \qquad (3.2.2)$$

where $\mathbf{G}$ is a full rank q×q matrix. The assumptions made on $\boldsymbol{\varepsilon}$ are

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}_{N\times 1} \text{ and } \mathrm{Var}[\boldsymbol{\varepsilon}] = \mathbf{R} \tag{3.2.3}$$

where $\mathbf{R}$ is a full rank N×N matrix. $\mathbf{U}$ and $\boldsymbol{\varepsilon}$ are assumed to be uncorrelated or

$$\mathrm{Cov}(\mathbf{U},\boldsymbol{\varepsilon}) = \mathbf{0}_{q\times N},$$

which implies

$$E\begin{bmatrix}\mathbf{U}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix} \text{ and } \mathrm{Var}\begin{bmatrix}\mathbf{U}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}_{q\times N}\\ \mathbf{0}_{N\times q} & \mathbf{R}\end{bmatrix}.$$

Therefore, using the expectation and variance of $\mathbf{U}$ and $\boldsymbol{\varepsilon}$, the expected value and variance of $\mathbf{y}$ are

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta} \text{ and } \mathrm{Var}[\mathbf{y}] = \mathbf{ZGZ}' + \mathbf{R} = \boldsymbol{\Sigma}. \tag{3.2.4}$$

One can see that when $\mathbf{R}=\sigma^2\mathbf{I}$ and $\mathbf{Z}=\mathbf{0}$, the mixed linear model reduces to the standard general linear model as given in equation 3.1.1.

3.3 Estimation in the Mixed Linear Model

Estimation in the mixed model setting is more complicated than in the general linear model. Besides the unknown parameters in $\boldsymbol{\beta}$, there are unknown parameters in $\mathbf{U}$, $\mathbf{G}$, and $\mathbf{R}$. Therefore, the usual least squares is not the best method for estimation.

Generalized least squares (GLS) may be used to obtain the best linear unbiased estimator of $\boldsymbol{\beta}$. GLS minimizes

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \tag{3.3.1}$$

where $\boldsymbol{\Sigma}$ is the variance of $\mathbf{y}$. However, knowledge of $\mathbf{G}$ and $\mathbf{R}$ is needed. Without known $\mathbf{G}$ and $\mathbf{R}$, an estimated GLS approach may be employed by using estimates for $\mathbf{G}$ and $\mathbf{R}$.

The most common approach for finding estimates for $\mathbf{G}$ and $\mathbf{R}$ are likelihood based methods. Under the assumption that $\mathbf{U}$ and $\boldsymbol{\varepsilon}$ are normally distributed, two likelihood based methods may be employed: maximum likelihood and restricted maximum likelihood. These methods will be discussed later in Section 3.5.

The normal equations from the generalized least squares in equation 3.3.1 are

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

and have solution

$$\mathbf{b} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}. \qquad (3.3.2)$$

In equation 3.3.2, $\boldsymbol{\Sigma}^{-1}$ or $\hat{\boldsymbol{\Sigma}}^{-1}$, depending on whether $\boldsymbol{\Sigma}$ is known or not, must be calculated. The matrix, $\boldsymbol{\Sigma}^{-1}$, has order equal to the total number of observations, which at times can be very large. Therefore, calculating $\boldsymbol{\Sigma}^{-1}$ is not an easy task. Henderson et al. (1959), however, showed that a set of equations that do not involve $\boldsymbol{\Sigma}^{-1}$ can be established. These equations are now known as Henderson's mixed model equations.

Henderson et al. (1959) showed that the mixed model equations are found by maximizing the joint density function of $\mathbf{y}$ and $\mathbf{U}$ with respect to $\boldsymbol{\beta}$ and $\mathbf{U}$. Under the assumption that $\mathbf{U}$ and $\boldsymbol{\varepsilon}$ are normally distributed i.e., $\mathbf{U} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$ and $\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$, the joint density function of $\mathbf{y}$ and $\mathbf{U}$ is:

$$f(\mathbf{y}, \mathbf{U}) = g(\mathbf{y} \mid \mathbf{U})\, h(\mathbf{U})$$
$$= C * \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{U})' \mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{U}) - \frac{1}{2}\mathbf{U}'\mathbf{G}^{-1}\mathbf{U}\right] \quad (3.3.3)$$

where C is a constant. Differentiating the natural logarithm of equation 3.3.3 with respect to $\boldsymbol{\beta}$ and U, the resulting derivatives are

$$\frac{\partial \ln f(\mathbf{y}, \mathbf{U})}{\partial \boldsymbol{\beta}} = -\frac{1}{2}\left[-2\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} + 2\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} + 2\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{U}\right]$$

and

$$\frac{\partial \ln f(\mathbf{y}, \mathbf{U})}{\partial \mathbf{U}} = -\frac{1}{2}\left[-2\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + 2\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} + 2\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{U} + 2\mathbf{G}^{-1}\mathbf{U}\right].$$

Equating the derivatives to zero yields the equations,

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{U}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$$
$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)\hat{\mathbf{U}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$$

which can be written in matrix form as

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{U}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \quad (3.3.4)$$

Once the estimates, $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, are found, Henderson's mixed model equations (3.3.4) are solved to get estimates of $\boldsymbol{\beta}$ and U. These estimates can be written as:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y} \tag{3.3.5}$$

$$\hat{\mathbf{U}} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{3.3.6}$$

where $\hat{\boldsymbol{\Sigma}} = \mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}' + \hat{\mathbf{R}}$. The solution of $\hat{\boldsymbol{\beta}}$ from Henderson's mixed model equations shown in equation 3.3.5 is equivalent to generalized least squares estimate given in equation 3.3.2.

When $\mathbf{G}$ and $\mathbf{R}$ are known, $\hat{\boldsymbol{\beta}}$ given in 3.3.5 is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and $\hat{\mathbf{U}}$ given in 3.3.6 is the best linear unbiased predictor (BLUP) of $\mathbf{U}$. If $\mathbf{G}$ and $\mathbf{R}$ are estimated, then $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{U}}$ are no longer BLUE and BLUP, respectively. The word empirical is often added to indicate this approximation. Therefore, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{U}}$ are then denoted as EBLUE and EBLUP (Littell et al., 1996). However, as $\hat{\boldsymbol{\Sigma}}$ approaches (in probability) $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\beta}}$ approaches (in probability) BLUE of $\boldsymbol{\beta}$ and $\hat{\mathbf{U}}$ approaches (in probability) BLUP of $\mathbf{U}$ (Henderson, 1984). The term BLUP for the random effects will be discussed in more detail following a discussion of the BLUE estimation of the fixed effects, $\boldsymbol{\beta}$.

The BLUE is the estimator of all linear unbiased estimators which has the minimum sampling variance. Suppose one is interested in the estimable function $\mathbf{k}'\boldsymbol{\beta}$. The BLUE of $\mathbf{k}'\boldsymbol{\beta}$ is $\mathbf{k}'\hat{\boldsymbol{\beta}}$ and the sampling variance of $\mathbf{k}'\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}
\mathrm{Var}\left[\mathbf{k}'\hat{\boldsymbol{\beta}}\right] &= \mathrm{Var}\left[\mathbf{k}'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}\right] \\
&= \mathbf{k}'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{k} \\
&= \mathbf{k}'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{k} \\
&= \mathbf{k}'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{k}.
\end{aligned}$$

Let $\mathbf{C}$ be a g-inverse of the coefficient matrix of the mixed model equations (3.3.4). Then $\mathbf{C}$ is written as

$$\mathbf{C}=\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}^{-}=\begin{bmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{bmatrix} \tag{3.3.7}$$

where, as given in McLean and Sanders (1988),

$$\begin{aligned}\mathbf{C}_{11}&=\left[\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}-\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\right)\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\right]^{-}\\ &=\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\end{aligned} \tag{3.3.8}$$

$$\mathbf{C}_{12}=-\mathbf{C}_{11}\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\right)\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1} \tag{3.3.9}$$

$$\mathbf{C}_{21}=\mathbf{C}_{12}' \tag{3.3.10}$$

and

$$\mathbf{C}_{22}=\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1}-\mathbf{C}_{21}\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\right)\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1}. \tag{3.3.11}$$

Using equation 3.3.8, the variance of $\mathbf{k}'\hat{\boldsymbol{\beta}}$ can be written as

$$\begin{aligned}\mathrm{Var}\left[\mathbf{k}'\hat{\boldsymbol{\beta}}\right]&=\mathbf{k}'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-}\mathbf{k}\\ &=\mathbf{k}'\mathbf{C}_{11}\mathbf{k}\end{aligned}$$

where $\mathbf{C}_{11}$ is the p×p upper submatrix of $\mathbf{C}$.

Next, the prediction of the random variables in $\mathbf{U}$ will be considered. For this situation, the prediction of a random variable can also be looked at as the estimation of the realized values of the random variables. This technique has been called BLUP by Henderson (1984). Robinson (1991) summarized the terminology of BLUP as follows:

BLUP estimates of the realized values of the random variables, **U** are:

1) Linear in the sense that they are linear functions of the data, **y**;
2) Unbiased in the sense that the average value of the estimate is equal to the average value of the quantity being estimated;
3) Best in the sense that they have minimum mean square error within the class of linear unbiased estimators; and
4) Predictors to distinguish them from estimators of fixed effects.

It is also of interest to note that $\hat{\mathbf{U}}$, the BLUP of **U**, is also an estimator of the conditional mean of **U** given **y**. From 3.3.6, $\hat{\mathbf{U}} = \mathbf{GZ}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$. From 3.2.1 and 3.2.2, it can be seen that $\mathrm{Cov}(\mathbf{U}, \mathbf{y}') = \mathbf{GZ}'$. Then, assuming normality and using 3.2.2 and 3.2.4,

$$\begin{aligned} \mathrm{E}[\mathbf{U} \mid \mathbf{y}] &= \mathrm{E}[\mathbf{U}] + \mathrm{Cov}(\mathbf{U}, \mathbf{y}')\left[\mathrm{Var}(\mathbf{y})\right]^{-1}\left[\mathbf{y} - \mathrm{E}[\mathbf{y}]\right] \\ &= \mathbf{0} + (\mathbf{GZ}')(\boldsymbol{\Sigma})^{-1}[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] \\ &= \mathbf{GZ}'\boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]. \end{aligned}$$

Thus, $\hat{\mathrm{E}}[\mathbf{U} \mid \mathbf{y}] = \mathbf{GZ}'\boldsymbol{\Sigma}^{-1}\left[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right] = \hat{\mathbf{U}}$. Finally, the following BLUP properties were derived by Henderson (1975).

1) $\hat{\mathbf{U}}$ is unique
2) $\mathrm{Cov}\left(\mathbf{k}'\hat{\boldsymbol{\beta}}, \hat{\mathbf{U}}'\right) = \mathbf{0}$
3) $\mathrm{Cov}\left(\mathbf{k}'\hat{\boldsymbol{\beta}}, \hat{\mathbf{U}}' - \mathbf{U}\right) = \mathbf{k}'\mathbf{C}_{12}$
4) $\mathrm{Var}\left(\hat{\mathbf{U}}\right) = \mathrm{Cov}\left(\hat{\mathbf{U}}, \mathbf{U}\right) = \mathbf{G} - \mathbf{C}_{22}$
5) $\mathrm{Var}\left(\hat{\mathbf{U}} - \mathbf{U}\right) = \mathbf{C}_{22}$

3.4 Inference in the Mixed Linear Model

The covariance matrix of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{U}}$ is the g-inverse of the coefficient matrix of the mixed model equations which is given in equation 3.3.7. But when $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$ are substituted into $\mathbf{C}$ in equation 3.3.7 to obtain $\hat{\mathbf{C}}$, everything becomes approximate. Consequently, hypothesis tests of interest are based on asymptotic t- and F-distributions.

For inference, estimable linear combinations of the form

$$\mathbf{L}'\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{U}\end{bmatrix}$$

are considered. Only the estimability of $\boldsymbol{\beta}$ is of concern since any linear combination of $\mathbf{U}$, the random effects, is estimable. For any estimable $\mathbf{L}$ matrix, $\mathbf{L}'\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{U}\end{bmatrix}$ is estimated by $\mathbf{L}'\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{U}}\end{bmatrix}$ and its approximate variance is $\mathbf{L}'\hat{\mathbf{C}}\mathbf{L}$.

If $\mathbf{L}$ is of rank 1, then the t-statistic for testing the hypothesis

$$H_0 : \mathbf{L}'\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{U}\end{bmatrix} = \mathbf{0} \tag{3.4.1}$$

is of the form

$$t = \frac{\mathbf{L}'\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{U}}\end{bmatrix}}{\sqrt{\mathbf{L}'\hat{\mathbf{C}}\mathbf{L}}}. \tag{3.4.2}$$

Under the assumed normality of $\mathbf{U}$ and $\boldsymbol{\varepsilon}$, t has an approximate t-distribution whose degrees of freedom must be approximated. The degrees of freedom are denoted as $\hat{\upsilon}$.

If $\mathbf{L}$ has of rank greater than 1, then a F statistic of the form

$$F = \frac{\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{U}}\end{bmatrix}' \mathbf{L}\left(\mathbf{L}'\hat{\mathbf{C}}\mathbf{L}\right)^{-1}\mathbf{L}'\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{U}}\end{bmatrix}}{\text{rank}(\mathbf{L})} \qquad (3.4.3)$$

is constructed for testing the hypothesis in 3.4.1. This F statistic has an approximate F-distribution with rank($\mathbf{L}$) numerator degrees of freedom and $\hat{v}$ denominator degrees of freedom.

A simple approach to approximating the degrees of freedom for the above test statistics is called the containment method. This method searches the random effect list for the effects that contain the fixed effect of interest. Among those, the rank contribution to the $[\mathbf{X} \quad \mathbf{Z}]$ matrix is computed and $\hat{v}$ is the smallest of the rank contributions. If no effects are found, $\hat{v}$ is approximated by $N - \text{rank}[\mathbf{X} \quad \mathbf{Z}]$. More advanced and computationally intensive methods are discussed in McLean and Sanders (1988). In these methods, the denominator degrees of freedom are approximated using the Satterthwaite procedure as discussed by Jeske and Harville (1988). Using the Jeske and Harville procedure, McLean and Sanders found that in the unbalanced case the denominator degrees of freedom decrease about one for each missing cell.

These test statistics treat $\hat{\mathbf{C}}$ as if it was the true $\mathbf{C}$. Therefore, inference may not be highly accurate. Kacker and Harville (1984) showed that the prediction error variance, $\mathbf{L}'\hat{\mathbf{C}}\mathbf{L}$, tends to be under estimated when the estimates, $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, made from the data are substituted into $\mathbf{L}'\mathbf{C}\mathbf{L}$. They propose the use of a correction term that decreases the bias in the estimated variance of $\mathbf{L}'\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{U}\end{bmatrix}$. McLean and Sanders (1988) showed that additional inflation of the variances may be required for inferences to be more accurate. The effect of using the correction to inflate the variance is noticeable when inference about specific

random effects is of interest. However, if inference about the random effects is not of interest and emphasis is placed only on the fixed effects, then the substitution of $\hat{\mathbf{C}}$ for $\mathbf{C}$ has little effect on the accuracy of the inference. If the substitution of $\hat{\mathbf{C}}$ for $\mathbf{C}$ causes unreliable or biased results, Wolfinger (1992) suggests a strategy of using smaller values of $\hat{\upsilon}$ to be more conservative.

When testing hypotheses regarding the mixed model, inference on the random effects may not be important. Many times the random effects are treated as nuisance parameters and included in the model chiefly for the reduction of the error term (Hicks, 1973). Therefore, it is not recommended to treat the random effects as fixed effects because the resulting analysis is likely to underestimate the variances of the estimated fixed effect means (Hsuan, 1993).

3.5 Variance Components

The mixed model allows for not only the modelling of the mean of $\mathbf{y}$ but also the variance of $\mathbf{y}$. The modelling of the variance is accomplished by specifying the structure of $\mathbf{G}$ and $\mathbf{R}$. The flexibility in the specification of $\mathbf{G}$ and $\mathbf{R}$ has changed over the years in mixed model theory. In the model specified by Hartley and Rao (1967), the variance of $\mathbf{U}$ was given by $\sigma^2\mathbf{I}$ and the variance of $\boldsymbol{\varepsilon}$ was given by $\sigma_\varepsilon^2\mathbf{I}$. In 1978, Harville assumed that the variance of $\boldsymbol{\varepsilon}$ was given by $\sigma_\varepsilon^2\mathbf{I}$ and that the variance of $\mathbf{U}$ was given by $\sigma_\varepsilon^2\mathbf{D}$ where $\mathbf{D}$ is a symmetric nonnegative definite matrix. The model specified by Jennrich and Schlucter in 1986 allowed $\mathbf{G}$, the variance of $\mathbf{U}$, to take on any structure. However, $\mathbf{R}$, the variance of $\boldsymbol{\varepsilon}$, was still specified by $\sigma^2\mathbf{I}$. Before this in 1982, Laird and Ware were some of the earliest to consider a practical application of the general formulation of $\mathbf{G}$ and $\mathbf{R}$. Although they stated that $\mathbf{R}$ could take on any form, their applications still set $\mathbf{R}=\sigma^2\mathbf{I}$ and used the $\mathbf{G}$ matrix to model variability.

The classical mixed model is defined with $\mathbf{R}=\sigma^2\mathbf{I}$ and $\mathbf{G}$ being a diagonal matrix containing variance components (Wolfinger, 1993). This model and all of the above model specifications are special cases of the general mixed model that permits arbitrary parameterized covariance structures in both $\mathbf{G}$ and $\mathbf{R}$. Defined below, for a 4x4 matrix, are some of the possible variance structures for $\mathbf{G}$ and $\mathbf{R}$.

Simple

$$\sigma^2\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

Compound symmetry

$$\begin{bmatrix}\sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2\end{bmatrix}$$

Unstructured

$$\begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14}\\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24}\\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34}\\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}\end{bmatrix}$$

Autoregressive

$$\sigma^2\begin{bmatrix}1 & \rho & \rho^2 & \rho^3\\ \rho & 1 & \rho & \rho^2\\ \rho^2 & \rho & 1 & \rho\\ \rho^3 & \rho^2 & \rho & 1\end{bmatrix}$$

Toeplitz

$$\begin{bmatrix}\sigma_1 & \sigma_2 & \sigma_3 & \sigma_4\\ \sigma_2 & \sigma_1 & \sigma_2 & \sigma_3\\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma_2\\ \sigma_4 & \sigma_3 & \sigma_2 & \sigma_1\end{bmatrix}$$

As indicated in Section 3.3, **G** and **R** are not usually known and must be estimated. The two likelihood methods used most often in the estimation of variance components are maximum likelihood (ML) and restricted maximum likelihood (REML). Hartley and Rao (1967) initially developed the procedure for the maximum likelihood estimation of the unknown constants and variances included in the general mixed analysis of variance model. Patterson and Thompson (1971) proposed a general from of restricted maximum likelihood estimation. The likelihoods that are to be maximized for each of these methods can be reduced to problems including the parameters in **G** and **R** only. The log-likelihoods can be written as follows:

$$\text{ML}: \quad ll(\mathbf{G}, \mathbf{R}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\log\left(d'\boldsymbol{\Sigma}^{-1}d\right) - \frac{N}{2}\left[1 + \log(2\pi/N)\right]$$

$$\text{REML}: \quad ll_{\text{REML}}(\mathbf{G}, \mathbf{R}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\log\left|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right| - \frac{N-p}{2}\log\left(d'\boldsymbol{\Sigma}^{-1}d\right) - \frac{N-p}{2}\left\{1 + \log\left[2\pi/(N-p)\right]\right\}$$

where $d = \mathbf{y} - \mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$ and p is the rank of **X**. Because closed form solutions for **G** and **R** do not exist, the likelihoods can be maximized over the unknown parameters in **G** and **R** and solved iteratively using a Newton-Raphson procedure to find optimal $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$. In using the Newton-Raphson procedure, the following system of equations are solved

$$\frac{\partial ll}{\partial \boldsymbol{\gamma}} = D(\boldsymbol{\gamma}) = \mathbf{0}$$

where $\boldsymbol{\gamma}$ is the vector of unknown parameters in $\mathbf{G}$ and $\mathbf{R}$. The algorithm for the Newton-Raphson consists of the following steps.

Step 1: Let $\boldsymbol{\gamma}^{\circ}$ be an initial estimate of $\boldsymbol{\gamma}$.

Step 2: Compute both $D(\boldsymbol{\gamma}^{\circ})$ which is $\frac{\partial ll}{\partial \boldsymbol{\gamma}}$ evaluated at $\boldsymbol{\gamma}^{\circ}$ and

$\mathbf{V}_{\gamma^{\circ}} = \mathbf{I}^{-1}(\boldsymbol{\gamma}^{\circ})$ which is the inverse of the information matrix evaluated at $\boldsymbol{\gamma}^{\circ}$.

Step 3: The refined solution then becomes $\boldsymbol{\gamma}^{1} = \boldsymbol{\gamma}^{\circ} + \mathbf{V}_{\gamma^{\circ}} D(\boldsymbol{\gamma}^{\circ})$.

Step 4: If $\|\boldsymbol{\gamma}^{\circ} - \boldsymbol{\gamma}^{1}\| < \Delta$ where Δ is a small positive number, then stop. Otherwise, set $\boldsymbol{\gamma}^{\circ} = \boldsymbol{\gamma}^{1}$ and repeat steps 2, 3, and 4.

When $\|\boldsymbol{\gamma}^{\circ} - \boldsymbol{\gamma}^{1}\| < \Delta$ is satisfied, then $\boldsymbol{\gamma}^{1}$ can be used as the optimal values for the parameters in $\mathbf{G}$ and $\mathbf{R}$ and lead to the optimal $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$.

The choice of ML or REML is determined by one's preference. Both method's have their advantages and disadvantages. One problem with both methods is that they are derived under the assumption of normality. Harville (1977), however, argued that the maximum likelihood estimators derived on the basis of normality may be suitable even when the form of the distribution is not specified. A problem with ML for the estimation of variance components is that it does not take account of the degrees of freedom used for estimating the model's fixed effects, whereas, REML does. ML, however, has the merit of simultaneously providing ML estimators of both fixed effects and variance components. On the other hand, REML variance component estimates are unaffected by the fixed effects. Finally, both ML and REML can be used for estimating variance components from unbalanced data with any mixed model.

3.6 Applications of the Mixed Linear Model

Mixed linear models serve as the basis for a variety of testing and estimation procedures. These procedures have been applied in many types of data including biological and agricultural data. The mixed model simplifies and unifies many common statistical analyses. It has been applied in many areas including repeated measures, random effects, split plot designs, random coefficients, and heterogeneous variances (Wolfinger, 1992).

For example, the mixed model is useful for analyzing repeated measures or longitudinal data. Given the general structure for the mixed model, the **R** matrix becomes an ideal place to model the covariance structure of the correlated data within a subject. There are two differences, however, between analyzing repeated measures data with the general linear model and the mixed model. When using the mixed model, all of the data that is known is used in constructing the likelihood. In contrast, the traditional multivariate general linear model works only with balanced data. In this model, subjects with missing data are deleted from the construction of the likelihood. Besides allowing subjects to have an unequal number of observation, the mixed model also allows for these observations to be taken at different time points for different subjects (Sherrull et al., 1994).

The second difference is in hypothesis testing. With the general linear model, two sets of tests of within subject effects can be produced, multivariate and univariate. The results of the multivariate F- tests are similar to those performed in the mixed model where **R** is an unstructured covariance matrix. These tests differ by a constant multiplier and may have different degrees of freedom. The univariate F tests are the same as the mixed model with a compound symmetry covariance structure. Again, the degrees of freedom may differ. These differences in degrees of freedom are not very critical since the F-statistics for the mixed model only have approximate F-distributions (Wolfinger, 1992).

The mixed model is not limited to unstructured or compound symmetry covariance structures when analyzing repeated measures data. Therefore, the mixed model may be preferred over the general linear model because of its flexibility with defining any covariance structure. The mixed model is highly advantageous especially when the number of time points becomes numerous. Multivariate models use fully parameterized covariance matrices and as the number of time points increase, the number of covariance parameters increase. As the covariance matrix becomes large, many of the covariance parameters become poorly estimated. Therefore, the covariance matrix can be modelled with the mixed model assuming various other structures that require fewer estimated parameters.

Grady and Helms (1995) discuss model selection techniques of the covariance matrix for incomplete longitudinal data. They discuss the basic structural covariance matrices: compound symmetry, autoregressive, and unstructured used for longitudinal studies, along with more complicated extensions of these covariances. When choosing the best covariance structure, they claim that it is up to the investigator to decide whether fitting models other than those provided automatically by standard software is worth the effort or if the basic covariance structures are sufficient for their data. If interest lies mostly on the fixed effects part of the model, then a basic covariance may be sufficient. Choosing a basic covariance is also sufficient if the choice of covariance structure has little effect on the fixed effects. However, if there is interest in the structure of the covariance matrix and/or the dependence of the responses over time, the alternative models offer information about the covariance that is not always available from the basic structural models. Laird and Ware (1982) also discuss the use of various covariance matrices for analyzing unbalanced repeated measures data.

The mixed model also allows for the inclusion of any random effects of interest. By including a random effect for subject, the correlation due to repeated observations on the same subject is taken into account, and thereby modelling the covariance matrix leads to

a better understanding of the data. In a study by Drum and McCullagh (1993), it was found that the inclusion of a random subject effect reduced the variability of the fixed effects by attributing that variability to the variability between subjects. The use of longitudinal random effects models is also discussed further in Jennrich and Schlucter (1986).

In the mixed model, the fixed effects provide estimates of the average responses in the population while the random effects account for the natural heterogeneity in the responses of different individuals and allow for the estimation of responses for each individual in the study (Pearson et al., 1994). Thus, in repeated measures analyses when there is a significant natural heterogeneity in the population and when individual level estimates are of interest, the mixed model may be preferred over the general linear model.

Pearson et al. (1994) describe several analyses that would not have been possible without the advances in longitudinal statistical methodology by using the mixed linear model. Since there are numerous types of biomedical research that involve the analysis of repeated measures, they state that the mixed model will become "an increasingly valuable tool in the studies because of their ability to test hypotheses, describe population average responses, and to provide individual-level responses". They also claim that the flexibility of the mixed effects regression models will make an important vehicle for advancing knowledge of the natural history of aging and disease.

For the mixed model, $\sigma^2\mathbf{I}$ is the most common form of $\mathbf{R}$ assumed. These cases include the randomized block design, split plot design, and random coefficients analysis. However, there are combination mixed models where both $\mathbf{G}$ and $\mathbf{R}$ are modelled non-trivially. These models are appropriate with time-series cross sectional data and multivariate repeated measures data. The combination mixed model is also appropriate in the case where $\mathbf{R}$ is diagonal but not constant i.e., a heterogeneous variance model. The simplest example of a heterogeneous variance model is when the different variances are

classified into groups. Wolfinger (1992) discusses more detailed examples of heterogeneous variance models.

Multivariate mixed models are also used for analyzing data. Henderson (1984), Meyer (1985), and Schaeffer and Wilton (1978) discuss multi-trait mixed models which can be considered multivariate mixed models since there is more than one response variable. In the multivariate setting, the **R** matrix handles the various covariances and variances for the various response variables. The **G** matrix is structured to handle the relationship among the random effects. The modelling of **G** and **R** leads to a strong parametric foundation (McLean, Sanders, and Stroup, 1991). Strong parametric procedures require fewer sampling units than conventional multivariate procedures. Also, as before with general univariate models, most multivariate procedures require complete records of data, whereas incomplete records can be used in the multivariate mixed model procedures.

Chapter 4
Power for the Mixed Linear Model

4.1 Introduction to Mixed Linear Model Power

Recall from Section 3.2 that the mixed linear model can be written in the following form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon} = [\mathbf{X} \quad \mathbf{Z}]\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{U}\end{bmatrix} + \boldsymbol{\varepsilon} \qquad (4.1.1)$$

where

$\mathbf{y}$ is a N×1 vector of measured responses,

$\mathbf{X}$ is a N×p known design matrix for the fixed effects with rank $\mathbf{X}$ = p where $p \le N$,

$\boldsymbol{\beta}$ is a p×1 vector of the unknown fixed effect parameters,

$\mathbf{Z}$ is a N×q known design matrix for the random effects,

$\mathbf{U}$ is a q×1 vector of unknown random effects,

and

$\boldsymbol{\varepsilon}$ is a N×1 vector of random errors.

Remember also that $\mathrm{Var}[\mathbf{U}] = \mathbf{G}_{q\times q}$ and $\mathrm{Var}[\boldsymbol{\varepsilon}] = \mathbf{R}_{N\times N}$. Once the estimates, $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, are found, the estimates of $\boldsymbol{\beta}$ and $\mathbf{U}$ are

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{y}$$
$$\hat{\mathbf{U}} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

where $\hat{\boldsymbol{\Sigma}} = \mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}' + \hat{\mathbf{R}}$.

In Section 3.4, the F statistic for testing the null hypothesis

$$H_0 : \mathbf{L}'\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{U}\end{bmatrix} = \mathbf{0}$$

was shown to be

$$F = \frac{\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{U}}\end{bmatrix}' \mathbf{L}\left(\mathbf{L}'\hat{\mathbf{C}}\mathbf{L}\right)^{-1}\mathbf{L}'\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{U}}\end{bmatrix}}{\text{rank}(\mathbf{L})}.$$

Suppose now it is of interest only to test the fixed effects portion of the mixed linear model. The null hypothesis becomes

$$H_0 : \mathbf{K}'\boldsymbol{\beta} = \mathbf{0}.$$

The matrix, **K**, contains the contrasts of the fixed effects that are of interest. When **K** has a rank greater than 1, then the test statistic associated with $H_1 : \mathbf{K}'\boldsymbol{\beta} \neq \mathbf{0}$ is

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{K}\left(\mathbf{K}'\hat{\mathbf{C}}_{11}\mathbf{K}\right)^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{\text{rank}(\mathbf{K})} \qquad (4.1.2)$$

where

$$\hat{\mathbf{C}}_{11} = \left[\mathbf{X}'\left(\mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}' + \hat{\mathbf{R}}\right)^{-1}\mathbf{X}\right]^{-1} .$$

This F statistic has an approximate F-distribution with rank($\mathbf{K}$) numerator degrees of freedom and $\hat{\upsilon}$ denominator degrees of freedom. The denominator degrees of freedom may be estimated by the same methods described in Section 3.4.

Using the same reasoning as Muller and Peterson (1984) did for noncentral F approximations of the central F approximations in the multivariate case, the test statistic given in 4.1.2 follows an approximate noncentral F-distribution with numerator degrees of freedom, rank($\mathbf{K}$), and denominator degrees of freedom, $\hat{\upsilon}$, and noncentrality parameter, λ, under the alternative hypothesis. The noncentrality parameter, as shown in Muller and Peterson, takes on the form of λ=(df model)F_A where F_A is the F that would be observed if one obtained $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, $\hat{\mathbf{G}} = \mathbf{G}$, and $\hat{\mathbf{R}} = \mathbf{R}$. Therefore, the noncentrality parameter may be expressed as

$$\begin{aligned}\lambda &= \text{rank}(\mathbf{K})\cdot \frac{\boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\mathbf{X}'(\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})^{-1}\mathbf{X}\right]^{-1}\mathbf{K}\right\}\mathbf{K}'\boldsymbol{\beta}}{\text{rank}(\mathbf{K})} \\ &= \boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\mathbf{X}'(\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})^{-1}\mathbf{X}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta} \\ &= \boldsymbol{\beta}'\mathbf{K}(\mathbf{K}'\mathbf{C}_{11}\mathbf{K})^{-1}\mathbf{K}'\boldsymbol{\beta}\end{aligned} \qquad (4.1.3)$$

Simulations by Helms and McCarroll (1991) also show that the distribution of the test statistic given in 4.1.2 is closely approximated by the noncentral F distribution with rank($\mathbf{K}$) numerator degrees of freedom and $\hat{\upsilon}$=N-rank ($\mathbf{X}$ $\mathbf{Z}$) denominator degrees of freedom, and noncentrality parameter λ. Therefore, 4.1.2 is an appropriate test statistic for the hypotheses of the fixed effects. Since an accurate approximation of the statistic's small sample noncentral distribution has been shown, power calculations for the tests of fixed

effects are feasible and are based on the approximate noncentral distribution of the test statistic. Therefore, the power of a size α test for the test statistic based on its approximate noncentral distribution is then given by

$$\Pi(\alpha, \boldsymbol{\Sigma}, \boldsymbol{\beta}) = 1\text{-}F(F_{crit}, \lambda, \text{rank}(\mathbf{K}), \hat{\upsilon}) \qquad (4.1.4)$$

where $F_{crit}=F^{-1}(1\text{-}\alpha, \text{rank}(\mathbf{K}), \hat{\upsilon})$ and F(.) is the cdf of the noncentral F described above evaluated at F_{crit}.

In Section 2.3, it was discussed how the noncentrality parameter, for the noncentral F used in hypothesis testing for the univariate GLM, could be expressed in terms of its distinct components. O'Brien and Muller (1993) showed that when the noncentrality was expressed as its distinct components, the noncentrality parameter could be written as the product of the total sample size and a primary noncentrality parameter. This primary noncentrality parameter was found not to be based on the total sample size but only the design points to be used, the weights of those points, and the conjectured values for the unknown parameters. Thus, a computationally more efficient method of computing power was gained by using the primary noncentrality parameter.

As in the univariate GLM case, the noncentrality parameter, λ (4.1.3), can be expressed more efficiently by writing the $\mathbf{C}_{11}$ portion in terms of its distinct components. For this presentation, three cases will be considered: Case I assumes that the column space for the random effects is fixed on the addition of another observation, Case II allows for the column space for the random effects to vary on the addition of another observation, and Case III allows for the column space for the random effects to vary as in Case II but it also allows for the row space for the fixed and random effects to vary from subject to subject or "experimental unit". The motivation behind the discussion of the three cases lies in the fact that the form of the noncentrality parameter varies from case to case.

As an example of Case I, suppose a multi-center clinical trial in which multiple treatments are to be studied will be performed. Treatments are considered as fixed effects and centers are to be taken as random effects. The center by treatment interactions will also be of interest, and therefore, they will also be considered random effects since center is a random effect. This is an example of Case I because the number of centers (and consequently, the column space of **Z**) is fixed.

A situation in which Case II is applicable is when a balanced and complete longitudinal study that involves a fixed number of treatments is considered. Each treatment will be evaluated on multiple occasions and be considered fixed. Subject will be considered as the random effect. When computing power, the number of subjects are added incrementally until the desired power is obtained. Therefore, since subject is the random effect, the column space of **Z** increases with the addition of another subject.

An example of Case III would be a multiple year clinical trial in which enrollment is permitted through the entire length of the trial. In this trial, subjects are randomized to various treatment groups. Subjects enrolled toward the end of the trial will not have complete data. Each treatment, which is considered as the fixed effect, is evaluated at the time of enrollment and until completion or termination of the study. Therefore, the row space of both **X** and **Z** vary from subject to subject depending on when they were entered into the study. Subject will again be considered as the random effect and therefore, the column space of **Z** will increase as each subject enters the clinical trial.

4.2 Noncentrality Parameter for Case I

Let $[\mathbf{X}^{**}\ \mathbf{Z}]$ be the $N_e \times (p+q)$ essence fixed/random design matrix formed by assembling the N_e unique rows of $[\mathbf{X}\ \ \mathbf{Z}]$. That is, $[\mathbf{X}^{**}\ \mathbf{Z}]$ is the collection of unique design points for the proposed study. Suppose $[\mathbf{X}\ \ \mathbf{Z}]$ has Nw_j rows identical to the j^{th}

row of $[\ddot{\mathbf{X}}\ \ddot{\mathbf{Z}}]$, with $0 < w_j < 1$ and $\sum w_j = 1$. Allowing $\mathbf{W}$ to be the $N_e \times N_e$ diagonal matrix having elements w_j, $\mathbf{NW}$ holds the N_e sample sizes. Next, the essence matrix, $[\ddot{\mathbf{X}}\ \ddot{\mathbf{Z}}]$, is partitioned into the separate essence fixed and essence random design matrices, denoted $\ddot{\mathbf{X}}$ ($N_e \times p$) and $\ddot{\mathbf{Z}}$ ($N_e \times q$), respectively.

For Case I, it is assumed that the column space for the random effects is fixed. That is, the column space of $\mathbf{Z}$ does not depend on the addition of another observation. Therefore, the number of columns of $\mathbf{Z}$ is equal to the number of columns of $\ddot{\mathbf{Z}}$ which is equal to q. Note that these individual essence matrices are not necessarily composed of unique rows and can be written in the form

$$\ddot{\mathbf{X}} = \begin{bmatrix} \mathbf{X}^* \\ \mathbf{X}^* \\ \vdots \\ \mathbf{X}^* \end{bmatrix} = \mathbf{1}_m \otimes \mathbf{X}^* \qquad (4.2.1)$$

and

$$\ddot{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}^* & & & \\ & \mathbf{Z}^* & & \\ & & \ddots & \\ & & & \mathbf{Z}^* \end{bmatrix} = \mathbf{I}(m) \otimes \mathbf{Z}^*. \qquad (4.2.2)$$

In 4.2.1, $\mathbf{X}^*$ is a $r \times p$ unique matrix and in 4.2.2, $\mathbf{Z}^*$ is the corresponding $r \times c$ unique matrix. The matrices $\mathbf{X}^*$ and $\mathbf{Z}^*$ are replicated in a stack or block diagonal form, respectively, m times. The number of $\mathbf{X}^*$ and $\mathbf{Z}^*$, m, is determined by the number of levels of the main random effects that are common to all random effects. Therefore, $N_e = mr$ and $q = mc$.

Going back to the clinical trial example, suppose there are three centers which are considered random and two treatments which are considered fixed. Also, the random

treatment by center interactions are considered. The fixed effect parameters for treatment 1 and 2, respectively, are β_1 and β_2 and the random effect parameters for the j^{th} center and the ij^{th} treatment by center interactions (i=1,2 and j=1,2,3) are $u_1, u_{11}, u_{21}, u_2, u_{12}, u_{22}, u_3, u_{13}$, and u_{23}. The essence matrix, $[\mathbf{X}^{\bullet\bullet}\ \mathbf{Z}]$, is

$$\left[\begin{array}{cc|ccccccccc} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}\right].$$

$\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$ are then given by

$$\ddot{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \ddot{\mathbf{Z}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The number of unique $\mathbf{X}^*$ and $\mathbf{Z}^*$ matrices corresponds to the number of centers which is three. This is found by noting that the random effects are center and the center by treatment interactions. Center is the main random effect that is common to all random effects and there are three levels of center. Since $N_e = 6$ and q=9, r must be $2 = N_e/m$ and c is $3 = q/m$. Hence, $\mathbf{X}^*$ has dimension 2×2 and $\mathbf{Z}^*$ has dimension 2×3 and they are

$$\mathbf{X}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{Z}^* = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Therefore, the essence matrices are

$$\ddot{\mathbf{X}} = \mathbf{1}_3 \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \ddot{\mathbf{Z}} = \mathbf{I}(3) \otimes \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

An initial step in finding the alternative form of the Case I noncentrality parameter is to show the following relationships

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \mathrm{N}\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{X}} \tag{4.2.3}$$

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} = \mathrm{N}\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}} \tag{4.2.4}$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} = \mathrm{N}\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{X}} \tag{4.2.5}$$

and

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} = \mathrm{N}\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}\ . \tag{4.2.6}$$

In order to prove the above statements, notation and rules of Kronecker products must first be established. The overall fixed effects design matrix, $\mathbf{X}$, can be written in the following form

$$\mathbf{X}_{\mathrm{N}\times\mathrm{p}} = \begin{bmatrix} \mathbf{1}_{f_1} \otimes \mathbf{X}^* \\ \mathbf{1}_{f_2} \otimes \mathbf{X}^* \\ \vdots \\ \mathbf{1}_{f_m} \otimes \mathbf{X}^* \end{bmatrix} \tag{4.2.7}$$

and the overall random effects design matrix, $\mathbf{Z}$, can be written as

$$\mathbf{Z}_{N\times q} = \begin{bmatrix} \mathbf{1}_{f_1} \otimes \mathbf{Z}^* & & & \\ & \mathbf{1}_{f_2} \otimes \mathbf{Z}^* & & \\ & & \ddots & \\ & & & \mathbf{1}_{f_m} \otimes \mathbf{Z}^* \end{bmatrix}. \tag{4.2.8}$$

Let **G** have the form

$$\mathbf{G} = \mathbf{I}(m) \otimes \mathbf{G}^* \tag{4.2.9}$$

where $\mathbf{G}^*$ is a c×c matrix. The weight matrix, **W**, has the form

$$\mathbf{W} = \begin{bmatrix} w_1\mathbf{I}(r) & & & \\ & w_2\mathbf{I}(r) & & \\ & & \ddots & \\ & & & w_m\mathbf{I}(r) \end{bmatrix} = \begin{bmatrix} \frac{f_1}{N}\mathbf{I}(r) & & & \\ & \frac{f_2}{N}\mathbf{I}(r) & & \\ & & \ddots & \\ & & & \frac{f_m}{N}\mathbf{I}(r) \end{bmatrix}. \tag{4.2.10}$$

Also, let **R** be of the form

$$\mathbf{R} = \mathbf{I}(f) \otimes \mathbf{R}^* \tag{4.2.11}$$

where $\mathbf{R}^*$ is a $r \times r$ matrix and $f = f_1 + f_2 + \cdots + f_m$. The essence random error variance matrix can be written as

$$\ddot{\mathbf{R}} = \mathbf{I}(m) \otimes \mathbf{R}^*. \tag{4.2.12}$$

Without loss of generality, assuming $\mathbf{R}^{*-1}$ exists and using the theorem regarding the inverse of a Kronecker product that states $(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$, the inverses of $\mathbf{R}$ and $\ddot{\mathbf{R}}$ are

$$\mathbf{R}^{-1}=\left(\mathbf{I}(f)\otimes\mathbf{R}^{*}\right)^{-1}=\mathbf{I}(f)\otimes\mathbf{R}^{*-1} \tag{4.2.13}$$

and

$$\ddot{\mathbf{R}}^{-1}=\left(\mathbf{I}(m)\otimes\mathbf{R}^{*}\right)^{-1}=\mathbf{I}(m)\otimes\mathbf{R}^{*-1}. \tag{4.2.14}$$

Finally, using equations 4.2.7-4.2.14 and the preceding assumptions, the proofs of equations 4.2.3-4.2.6 are

Proof of Equality 4.2.3: $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}=\mathbf{N}\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{X}}$

$$\begin{aligned}
&\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\\
&=\begin{bmatrix}\left(\mathbf{1}_{f_1}\otimes\mathbf{X}^{*}\right)' & \left(\mathbf{1}_{f_2}\otimes\mathbf{X}^{*}\right)' & \cdots & \left(\mathbf{1}_{f_m}\otimes\mathbf{X}^{*}\right)'\end{bmatrix}
\begin{bmatrix}\mathbf{I}(f_1)\otimes\mathbf{R}^{*-1} & & & \\ & \mathbf{I}(f_2)\otimes\mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{I}(f_m)\otimes\mathbf{R}^{*-1}\end{bmatrix}\\
&\qquad\begin{bmatrix}\mathbf{1}_{f_1}\otimes\mathbf{X}^{*}\\ \mathbf{1}_{f_2}\otimes\mathbf{X}^{*}\\ \vdots\\ \mathbf{1}_{f_m}\otimes\mathbf{X}^{*}\end{bmatrix}\\
&=\left(\mathbf{1}_{f_1}\otimes\mathbf{X}^{*}\right)'\left(\mathbf{I}(f_1)\otimes\mathbf{R}^{*-1}\right)\left(\mathbf{1}_{f_1}\otimes\mathbf{X}^{*}\right)+\left(\mathbf{1}_{f_2}\otimes\mathbf{X}^{*}\right)'\left(\mathbf{I}(f_2)\otimes\mathbf{R}^{*-1}\right)\left(\mathbf{1}_{f_2}\otimes\mathbf{X}^{*}\right)+\cdots\\
&\qquad+\left(\mathbf{1}_{f_m}\otimes\mathbf{X}^{*}\right)'\left(\mathbf{I}(f_m)\otimes\mathbf{R}^{*-1}\right)\left(\mathbf{1}_{f_m}\otimes\mathbf{X}^{*}\right)\\
&=f_1\mathbf{X}^{*}{}'\mathbf{R}^{*-1}\mathbf{X}^{*}+f_2\mathbf{X}^{*}{}'\mathbf{R}^{*-1}\mathbf{X}^{*}+\cdots+f_m\mathbf{X}^{*}{}'\mathbf{R}^{*-1}\mathbf{X}^{*}
\end{aligned}$$

$$= N\left[\frac{f_1}{N}X^{*'}R^{*-1}X^{*} + \frac{f_2}{N}X^{*'}R^{*-1}X^{*} + \cdots + \frac{f_m}{N}X^{*'}R^{*-1}X^{*}\right]$$

$$= N\left[\frac{f_1}{N}X^{*'}R^{*-1} \quad \frac{f_2}{N}X^{*'}R^{*-1} \quad \cdots \quad \frac{f_m}{N}X^{*'}R^{*-1}\right]\begin{bmatrix}X^{*}\\X^{*}\\\vdots\\X^{*}\end{bmatrix}$$

$$= N\left[\frac{f_1}{N}X^{*'} \quad \frac{f_2}{N}X^{*'} \quad \cdots \quad \frac{f_m}{N}X^{*'}\right]\begin{bmatrix}R^{*-1} & & & \\ & R^{*-1} & & \\ & & \ddots & \\ & & & R^{*-1}\end{bmatrix}\begin{bmatrix}X^{*}\\X^{*}\\\vdots\\X^{*}\end{bmatrix}$$

$$= N\left[X^{*'} \quad X^{*'} \quad \cdots \quad X^{*'}\right]\begin{bmatrix}\frac{f_1}{N}I(r) & & & \\ & \frac{f_2}{N}I(r) & & \\ & & \ddots & \\ & & & \frac{f_m}{N}I(r)\end{bmatrix}\begin{bmatrix}R^{*-1} & & & \\ & R^{*-1} & & \\ & & \ddots & \\ & & & R^{*-1}\end{bmatrix}\begin{bmatrix}X^{*}\\X^{*}\\\vdots\\X^{*}\end{bmatrix}$$

$$= N\ddot{X}'W\ddot{R}^{-1}\ddot{X}$$

End of Proof for 4.2.3

Proof of Equality 4.2.4: $\mathbf{X'R^{-1}Z = N\ddot{X}'W\ddot{R}^{-1}\ddot{Z}}$

$$\mathbf{X'R^{-1}Z}$$

$$= \begin{bmatrix} \left(\mathbf{1}_{f_1} \otimes \mathbf{X}^*\right)' & \left(\mathbf{1}_{f_2} \otimes \mathbf{X}^*\right)' & \cdots & \left(\mathbf{1}_{f_m} \otimes \mathbf{X}^*\right)' \end{bmatrix} \begin{bmatrix} \mathbf{I}(f_1) \otimes \mathbf{R}^{*-1} & & & \\ & \mathbf{I}(f_2) \otimes \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{I}(f_m) \otimes \mathbf{R}^{*-1} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{1}_{f_1} \otimes \mathbf{Z}^* & & & \\ & \mathbf{1}_{f_2} \otimes \mathbf{Z}^* & & \\ & & \ddots & \\ & & & \mathbf{1}_{f_m} \otimes \mathbf{Z}^* \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_m \end{bmatrix} \qquad \text{where } \mathbf{A}_i = \left(\mathbf{1}_{f_i} \otimes \mathbf{X}^*\right)' \left(\mathbf{I}(f_i) \otimes \mathbf{R}^{*-1}\right)\left(\mathbf{1}_{f_i} \otimes \mathbf{Z}^*\right)$$

$$= \begin{bmatrix} f_1 \mathbf{X}^{*'} \mathbf{R}^{*-1} \mathbf{Z}^* & f_2 \mathbf{X}^{*'} \mathbf{R}^{*-1} \mathbf{Z}^* & \cdots & f_m \mathbf{X}^{*'} \mathbf{R}^{*-1} \mathbf{Z}^* \end{bmatrix}$$

$$= \begin{bmatrix} N\mathbf{X}^{*'} \frac{f_1}{N} \mathbf{R}^{*-1} \mathbf{Z}^* & N\mathbf{X}^{*'} \frac{f_2}{N} \mathbf{R}^{*-1} \mathbf{Z}^* & \cdots & N\mathbf{X}^{*'} \frac{f_m}{N} \mathbf{R}^{*-1} \mathbf{Z}^* \end{bmatrix}$$

$$= N \begin{bmatrix} \mathbf{X}^{*'} \frac{f_1}{N} \mathbf{R}^{*-1} & \mathbf{X}^{*'} \frac{f_2}{N} \mathbf{R}^{*-1} & \cdots & \mathbf{X}^{*'} \frac{f_m}{N} \mathbf{R}^{*-1} \end{bmatrix} \begin{bmatrix} \mathbf{Z}^* & & & \\ & \mathbf{Z}^* & & \\ & & \ddots & \\ & & & \mathbf{Z}^* \end{bmatrix}$$

$$= N \begin{bmatrix} \mathbf{X}^{*'} \frac{f_1}{N} & \mathbf{X}^{*'} \frac{f_2}{N} & \cdots & \mathbf{X}^{*'} \frac{f_m}{N} \end{bmatrix} \begin{bmatrix} \mathbf{R}^{*-1} & & & \\ & \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{R}^{*-1} \end{bmatrix} \begin{bmatrix} \mathbf{Z}^* & & & \\ & \mathbf{Z}^* & & \\ & & \ddots & \\ & & & \mathbf{Z}^* \end{bmatrix}$$

$$= N\begin{bmatrix} \mathbf{X}^{*\prime} & \mathbf{X}^{*\prime} & \cdots & \mathbf{X}^{*\prime} \end{bmatrix} \begin{bmatrix} \frac{f_1}{N}\mathbf{I}(r) & & & \\ & \frac{f_2}{N}\mathbf{I}(r) & & \\ & & \ddots & \\ & & & \frac{f_m}{N}\mathbf{I}(r) \end{bmatrix} \begin{bmatrix} \mathbf{R}^{*-1} & & & \\ & \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{R}^{*-1} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{Z}^{*} & & & \\ & \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*} \end{bmatrix}$$

$$= N\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}$$

End of Proof for 4.2.4

A similar proof can be shown for Equality 4.2.5 ($\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} = N\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{X}}$) and the proof of the final statement is given by

Proof of Equality 4.2.6: $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} = N\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$$

$$= \begin{bmatrix} \mathbf{1}_{f_1}' \otimes \mathbf{Z}^{*\prime} & & & \\ & \mathbf{1}_{f_2}' \otimes \mathbf{Z}^{*\prime} & & \\ & & \ddots & \\ & & & \mathbf{1}_{f_m}' \otimes \mathbf{Z}^{*\prime} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{I}(f_1) \otimes \mathbf{R}^{*-1} & & & \\ & \mathbf{I}(f_2) \otimes \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{I}(f_m) \otimes \mathbf{R}^{*-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}_{f_1} \otimes \mathbf{Z}^{*} & & & \\ & \mathbf{1}_{f_2} \otimes \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & \mathbf{1}_{f_m} \otimes \mathbf{Z}^{*} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \mathbf{A}_m \end{bmatrix} \quad \text{where } \mathbf{A}_i = \left(\mathbf{1}_{f_i}' \otimes \mathbf{Z}^{*'}\right)\left(\mathbf{I}(f_i) \otimes \mathbf{R}^{*-1}\right)\left(\mathbf{1}_{f_i} \otimes \mathbf{Z}^{*}\right)$$

$$= \begin{bmatrix} f_1 \mathbf{Z}^{*'} \mathbf{R}^{*-1} \mathbf{Z}^{*} & & & \\ & f_2 \mathbf{Z}^{*'} \mathbf{R}^{*-1} \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & f_m \mathbf{Z}^{*'} \mathbf{R}^{*-1} \mathbf{Z}^{*} \end{bmatrix}$$

$$= \begin{bmatrix} N\mathbf{Z}^{*'} \frac{f_1}{N} \mathbf{R}^{*-1} \mathbf{Z}^{*} & & & \\ & N\mathbf{Z}^{*'} \frac{f_2}{N} \mathbf{R}^{*-1} \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & N\mathbf{Z}^{*'} \frac{f_m}{N} \mathbf{R}^{*-1} \mathbf{Z}^{*} \end{bmatrix}$$

$$= N \begin{bmatrix} \mathbf{Z}^{*'} \frac{f_1}{N} \mathbf{R}^{*-1} & & & \\ & \mathbf{Z}^{*'} \frac{f_2}{N} \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*'} \frac{f_m}{N} \mathbf{R}^{*-1} \end{bmatrix} \begin{bmatrix} \mathbf{Z}^{*} & & & \\ & \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*} \end{bmatrix}$$

$$= N \begin{bmatrix} \mathbf{Z}^{*'} \frac{f_1}{N} & & & \\ & \mathbf{Z}^{*'} \frac{f_2}{N} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*'} \frac{f_m}{N} \end{bmatrix} \begin{bmatrix} \mathbf{R}^{*-1} & & & \\ & \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{R}^{*-1} \end{bmatrix} \begin{bmatrix} \mathbf{Z}^{*} & & & \\ & \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*} \end{bmatrix}$$

$$= N\begin{bmatrix} \mathbf{Z}^{*'} & & & \\ & \mathbf{Z}^{*'} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*'} \end{bmatrix}\begin{bmatrix} \frac{f_1}{N}\mathbf{I}(r) & & & \\ & \frac{f_2}{N}\mathbf{I}(r) & & \\ & & \ddots & \\ & & & \frac{f_m}{N}\mathbf{I}(r) \end{bmatrix}\begin{bmatrix} \mathbf{R}^{*-1} & & & \\ & \mathbf{R}^{*-1} & & \\ & & \ddots & \\ & & & \mathbf{R}^{*-1} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{Z}^{*} & & & \\ & \mathbf{Z}^{*} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{*} \end{bmatrix}$$

$$= N\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}$$

End of Proof for 4.2.6

Recall the goal has been to write the $\mathbf{C}_{11}$ portion of the noncentrality parameter in terms of its distinct components where $\mathbf{C}_{11}$ is given by $\left[\mathbf{X}'(\mathbf{ZGZ}'+\mathbf{R})^{-1}\mathbf{X}\right]^{-1}$. Given the proofs of statements 4.2.3-4.2.6 and invoking the equality stated in Rao (1973),

$$(\mathbf{A}+\mathbf{BDB}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}+\mathbf{D}^{-1}\right)^{-1}\mathbf{B}'\mathbf{A}^{-1},$$

the portion of $\mathbf{C}_{11}$ that is inverted is then

$$\mathbf{X}'(\mathbf{ZGZ}'+\mathbf{R})^{-1}\mathbf{X}$$
$$= \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} - \left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\right)\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}$$
$$= N\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{X}} - \left(N\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}\right)\left(N\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}+\mathbf{G}^{-1}\right)^{-1}N\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{X}}$$
$$= N\ddot{\mathbf{X}}'\mathbf{W}\left[\ddot{\mathbf{R}}^{-1} - N\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}\left(N\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\ddot{\mathbf{Z}}+\mathbf{G}^{-1}\right)^{-1}\ddot{\mathbf{Z}}'\mathbf{W}\ddot{\mathbf{R}}^{-1}\right]\ddot{\mathbf{X}}$$
$$= N\ddot{\mathbf{X}}'\mathbf{W}\left[\ddot{\mathbf{R}}^{-1}\mathbf{W}^{-1}\mathbf{W} - \ddot{\mathbf{R}}^{-1}\mathbf{W}^{-1}\mathbf{W}\sqrt{N}\ddot{\mathbf{Z}}\left(\ddot{\mathbf{Z}}'\sqrt{N}\mathbf{W}\ddot{\mathbf{R}}^{-1}\mathbf{W}^{-1}\mathbf{W}\sqrt{N}\ddot{\mathbf{Z}}+\mathbf{G}^{-1}\right)^{-1}\right.$$
$$\left.\ddot{\mathbf{Z}}'\sqrt{N}\mathbf{W}\ddot{\mathbf{R}}^{-1}\mathbf{W}^{-1}\mathbf{W}\right]\ddot{\mathbf{X}}$$

$$= N\ddot{X}'W\left\{\left(W\ddot{R}\right)^{-1} - \left(W\ddot{R}\right)^{-1}W\sqrt{N}\ddot{Z}\left[\ddot{Z}'\sqrt{N}W\left(W\ddot{R}\right)^{-1}W\sqrt{N}\ddot{Z} + G^{-1}\right]^{-1}\right.$$
$$\left.\ddot{Z}'\sqrt{N}W\left(W\ddot{R}\right)^{-1}\right\}W\ddot{X}$$
$$= N\ddot{X}'W\left\{W\ddot{R} + \left(W\sqrt{N}\ddot{Z}\right)G\left(\ddot{Z}'\sqrt{N}W\right)\right\}^{-1}W\ddot{X}$$
$$= N\ddot{X}'W\left\{W^{-1}\left[W\ddot{R} + \left(W\sqrt{N}\ddot{Z}\right)G\left(\ddot{Z}'\sqrt{N}W\right)\right]\right\}^{-1}\ddot{X}$$
$$= N\ddot{X}'W\left\{N\ddot{Z}G\ddot{Z}'W + \ddot{R}\right\}^{-1}\ddot{X} \qquad (4.2.15)$$

Therefore, using the equality in 4.2.15, the Case I noncentrality parameter is shown to be

$$\lambda = \beta'K\left(K'C_{11}K\right)^{-1}K'\beta$$
$$= \beta'K\left\{K'\left[X'\left(ZGZ' + R\right)^{-1}X\right]^{-1}K\right\}^{-1}K'\beta$$
$$= \beta'K\left\{K'\left[N\ddot{X}'W\left(N\ddot{Z}G\ddot{Z}'W + \ddot{R}\right)^{-1}\ddot{X}\right]^{-1}K\right\}^{-1}K'\beta$$
$$= N\beta'K\left\{K'\left[\ddot{X}'W\left(N\ddot{Z}G\ddot{Z}'W + \ddot{R}\right)^{-1}\ddot{X}\right]^{-1}K\right\}^{-1}K'\beta$$
$$= N\lambda_1^*$$

It was hoped that λ_1^*, the primary noncentrality for Case I, could be written so that it is not related to N. This will occur if (a) $\mathbf{Z=0}$ (i.e., no random effects) or if (b) $\mathbf{G=0}$, neither of which are interesting simplifications. For now, however, there still exists a more computationally efficient version of λ since matrices of smaller dimension are being inverted and all pieces of λ_1^* are fixed except N, the total sample size.

4.3 Noncentrality Parameter for Case II

Assume the case where the column space for the random effects is not fixed. This occurs when each subject has its own vector of random effects. This situation is common

to repeated measures models. The overall design matrices, **X** and **Z**, may be represented as

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{f_1} \otimes \mathbf{X}_1 \\ \mathbf{1}_{f_2} \otimes \mathbf{X}_2 \\ \vdots \\ \mathbf{1}_{f_T} \otimes \mathbf{X}_T \end{bmatrix} \tag{4.3.1}$$

and

$$\mathbf{Z} = \begin{bmatrix} \mathbf{I}(f_1) \otimes \mathbf{Z}^{\circ} & & & \\ & \mathbf{I}(f_2) \otimes \mathbf{Z}^{\circ} & & \\ & & \ddots & \\ & & & \mathbf{I}(f_T) \otimes \mathbf{Z}^{\circ} \end{bmatrix} \tag{4.3.2}$$

where $\mathbf{X}_t$ is a $r \times p$ matrix and $\mathbf{Z}^{\circ}$ is a $r \times c$ matrix. Also, each $\mathbf{X}_t$ is a unique fixed effects design matrix for a given subject and $\mathbf{Z}^{\circ}$ is the corresponding random effects design matrix. Notice that the corresponding matrix for each $\mathbf{X}_t$ is the same matrix, $\mathbf{Z}^{\circ}$. The design of the study specifies that the set of observations represented by $\mathbf{X}_t$ and $\mathbf{Z}^{\circ}$ are replicated f_t times. The essence matrices are

$$\ddot{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_T \end{bmatrix} \tag{4.3.3}$$

and

$$\ddot{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}^{\circ} & & & \\ & \mathbf{Z}^{\circ} & & \\ & & \ddots & \\ & & & \mathbf{Z}^{\circ} \end{bmatrix} = \mathbf{I}(T) \otimes \mathbf{Z}^{\circ} \tag{4.3.4}$$

where $\ddot{\mathbf{X}}$ is a $N_e \times p$ matrix and $\ddot{\mathbf{Z}}$ is a $N_e \times (cT)$ matrix. Note that the number of columns of $\mathbf{Z}$ do not equal the number of columns of $\ddot{\mathbf{Z}}$.

Returning to the longitudinal study example, suppose a two treatment study evaluated on three occasions is to be performed. It is assumed that the dependent variable of interest varies linearly with time and a separate regression will be fit for each treatment. Thus, the fixed effect parameters are given by $\beta_{01}, \beta_1, \beta_{02}$, and β_2 where β_{0i} and β_i are the intercept and slope parameters for the i^{th} treatment. Each subject has its own random effect parameter u_j. The three fixed time points are given by x_1, x_2, and x_3. The essence matrices are then

$$\ddot{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{bmatrix} = \left[\begin{array}{cccc} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ 1 & x_3 & 0 & 0 \\ \hline 0 & 0 & 1 & x_1 \\ 0 & 0 & 1 & x_2 \\ 0 & 0 & 1 & x_3 \end{array}\right]$$

and

$$\ddot{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}^* & \\ & \mathbf{Z}^* \end{bmatrix} = \left[\begin{array}{c:c} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hdashline 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{array}\right]$$

For Case II, the general form of $\mathbf{R}$ is

$$\mathbf{R} = \mathbf{I}(f) \otimes \mathbf{R}^* \qquad (4.3.5)$$

where $\mathbf{R}^*$ is a $r \times r$ matrix and $f = \sum_{i=1}^{T} f_i$. The essence matrix, $\ddot{\mathbf{R}}$, is

$$\ddot{\mathbf{R}} = \mathbf{I}(T) \otimes \mathbf{R}^*. \tag{4.3.6}$$

$\mathbf{G}$ has the form

$$\mathbf{G} = \mathbf{I}(f) \otimes \mathbf{G}^* \tag{4.3.7}$$

where $\mathbf{G}^*$ is a $c \times c$ matrix. The essence matrix, $\ddot{\mathbf{G}}$, is

$$\ddot{\mathbf{G}} = \mathbf{I}(T) \otimes \mathbf{G}^*. \tag{4.3.8}$$

The weight matrix, $\mathbf{W}$, has the form

$$\mathbf{W} = \begin{bmatrix} \frac{f_1}{N}\mathbf{I}(r) & & & \\ & \frac{f_2}{N}\mathbf{I}(r) & & \\ & & \ddots & \\ & & & \frac{f_T}{N}\mathbf{I}(r) \end{bmatrix} \tag{4.3.9}$$

Thus for Case II, it can shown, by using equations 4.3.1-4.3.9, that the inverted portion of $\mathbf{C}_{11}$ can be written in its distinct components as

$$\mathbf{X}'(\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})^{-1}\mathbf{X} = N\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}' + \ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}} \tag{4.3.10}$$

Proof of Equality 4.3.10

First, working with the inner piece of $\mathbf{X}'(\mathbf{ZGZ}'+\mathbf{R})^{-1}\mathbf{X}$ leads to

$$(\mathbf{ZGZ}'+\mathbf{R})$$

$$= \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{Z}^* & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{Z}^* \end{bmatrix} \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{G}^* & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{G}^* \end{bmatrix} \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{Z}^{*\prime} & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{Z}^{*\prime} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{R}^* & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{R}^* \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_T \end{bmatrix} \quad \text{where } \mathbf{A}_i = \left(\mathbf{I}(f_i)\otimes\mathbf{Z}^*\right)\left(\mathbf{I}(f_i)\otimes\mathbf{G}^*\right)\left(\mathbf{I}(f_i)\otimes\mathbf{Z}^{*\prime}\right)+\left(\mathbf{I}(f_i)\otimes\mathbf{R}^*\right)$$

$$= \begin{bmatrix} \left(\mathbf{I}(f_1)\otimes\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}\right)+\left(\mathbf{I}(f_1)\otimes\mathbf{R}^*\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}\right)+\left(\mathbf{I}(f_T)\otimes\mathbf{R}^*\right) \end{bmatrix}$$

$$= \begin{bmatrix} \left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}+\mathbf{R}^*\right)\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}+\mathbf{R}^*\right)\right) \end{bmatrix}$$

This piece inverted is then,

$$(\mathbf{ZGZ}'+\mathbf{R})^{-1}=\begin{bmatrix}\left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\right)\end{bmatrix}$$

So,

$$\mathbf{X}'(\mathbf{ZGZ}'+\mathbf{R})^{-1}\mathbf{X}$$

$$=\begin{bmatrix}\mathbf{1}'_{f_1}\otimes\mathbf{X}'_1 & \cdots & \mathbf{1}'_{f_T}\otimes\mathbf{X}'_T\end{bmatrix}\begin{bmatrix}\left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\right)\end{bmatrix}\begin{bmatrix}\mathbf{1}_{f_1}\otimes\mathbf{X}_1\\ \vdots\\ \mathbf{1}_{f_T}\otimes\mathbf{X}_T\end{bmatrix}$$

$$=\left(\mathbf{1}'_{f_1}\otimes\mathbf{X}'_1\right)\left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\right)\left(\mathbf{1}_{f_1}\otimes\mathbf{X}_1\right)+\cdots$$

$$+\left(\mathbf{1}'_{f_T}\otimes\mathbf{X}'_T\right)\left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\right)\left(\mathbf{1}_{f_T}\otimes\mathbf{X}_T\right)$$

$$=\left(\mathbf{1}'_{f_1}\mathbf{I}(f_1)\mathbf{1}_{f_1}\otimes\mathbf{X}'_1\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_1\right)+\cdots+\left(\mathbf{1}'_{f_T}\mathbf{I}(f_T)\mathbf{1}_{f_T}\otimes\mathbf{X}'_T\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_T\right)$$

$$=\left(f_1\otimes\mathbf{X}'_1\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_1\right)+\cdots+\left(f_T\otimes\mathbf{X}'_T\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_T\right)$$

$$=f_1\mathbf{X}'_1\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_1+\cdots+f_T\mathbf{X}'_T\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_T$$

$$=N\mathbf{X}'_1\frac{f_1}{N}\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_1+\cdots+N\mathbf{X}'_T\frac{f_T}{N}\left(\mathbf{Z}^{\circ}\mathbf{G}^{*}\mathbf{Z}^{\circ\prime}+\mathbf{R}^{*}\right)^{-1}\mathbf{X}_T$$

$$= N\begin{bmatrix}\mathbf{X}_1' & \cdots & \mathbf{X}_T'\end{bmatrix}\begin{bmatrix}\frac{f_1}{N}\mathbf{I}(r) & & \\ & \ddots & \\ & & \frac{f_T}{N}\mathbf{I}(r)\end{bmatrix}\begin{bmatrix}\left(\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}+\mathbf{R}^*\right)^{-1} & & \\ & \ddots & \\ & & \left(\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}+\mathbf{R}^*\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T\end{bmatrix}$$

$$= N\begin{bmatrix}\mathbf{X}_1' & \cdots & \mathbf{X}_T'\end{bmatrix}\begin{bmatrix}\frac{f_1}{N}\mathbf{I}(r) & & \\ & \ddots & \\ & & \frac{f_T}{N}\mathbf{I}(r)\end{bmatrix}\begin{bmatrix}\mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}+\mathbf{R}^* & & \\ & \ddots & \\ & & \mathbf{Z}^*\mathbf{G}^*\mathbf{Z}^{*\prime}+\mathbf{R}^*\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T\end{bmatrix}$$

$$= N\begin{bmatrix}\mathbf{X}_1' & \cdots & \mathbf{X}_T'\end{bmatrix}\begin{bmatrix}\frac{f_1}{N}\mathbf{I}(r) & & \\ & \ddots & \\ & & \frac{f_T}{N}\mathbf{I}(r)\end{bmatrix}$$

$$\left\{\begin{bmatrix}\mathbf{Z}^* & & \\ & \ddots & \\ & & \mathbf{Z}^*\end{bmatrix}\begin{bmatrix}\mathbf{G}^* & & \\ & \ddots & \\ & & \mathbf{G}^*\end{bmatrix}\begin{bmatrix}\mathbf{Z}^{*\prime} & & \\ & \ddots & \\ & & \mathbf{Z}^{*\prime}\end{bmatrix}+\begin{bmatrix}\mathbf{R}^* & & \\ & \ddots & \\ & & \mathbf{R}^*\end{bmatrix}\right\}^{-1}\begin{bmatrix}\mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T\end{bmatrix}$$

$$= N\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}'+\ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}$$

End of Proof for 4.3.10

Therefore, using the equality stated in 4.3.10, the noncentrality parameter for Case II is

$$\lambda = \boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\mathbf{X}'(\mathbf{Z}\mathbf{G}\mathbf{Z}'+\mathbf{R})^{-1}\mathbf{X}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}$$

$$= \boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[N\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}'+\ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}$$

$$= N\boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}'+\ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}$$

$$= N\lambda_{II}^*$$

Note here that λ_{II}^{*}, the primary noncentrality for Case II, is not related to N. It is also of importance to note that N stands for the total number of observed responses. Assuming a complete and balanced design, the total sample size necessary is determined by dividing N by the number of occasions per subject which is r.

4.4 Noncentrality Parameter for Case III

A third case is similar to the situation in Case II but it allows for each $\mathbf{X}_t$, and therefore, each corresponding $\mathbf{Z}_t$ (i=1...T), to have differing numbers of rows. This case has been described in Helms (1992); however, he assumes $\mathrm{Var}[\boldsymbol{\varepsilon}] = \mathbf{R} = \sigma^2\mathbf{I}$. For this presentation, $\mathbf{R}$ will be allowed to take on any block diagonal form.

The overall fixed effects design matrix, $\mathbf{X}$, and the essence matrix, $\ddot{\mathbf{X}}$, can be written as they are in Case II, equations 4.3.1 and 4.3.3, respectively, with the exception that the $\mathbf{X}_t$ (t=1...T) are $r_t \times p$ matrices. The overall random effect matrix, $\mathbf{Z}$, now has the form

$$\mathbf{Z} = \begin{bmatrix} \mathbf{I}(f_1) \otimes \mathbf{Z}_1 & & & \\ & \mathbf{I}(f_2) \otimes \mathbf{Z}_2 & & \\ & & \ddots & \\ & & & \mathbf{I}(f_T) \otimes \mathbf{Z}_T \end{bmatrix} \qquad (4.4.1)$$

where $\mathbf{Z}_t$ is a $r_t \times c$ matrix. Again, each $\mathbf{X}_t$ is a unique fixed effects design matrix for a given subject and $\mathbf{Z}_t$ is the corresponding random effects design matrix. However in this case, the $\mathbf{Z}_t$ matrices differ when $r_i \neq r_j$, and the $\mathbf{Z}_i$ matrices are subsets of the $\mathbf{Z}_j$ matrices when $r_j > r_i$. The design of the study again specifies that the set of observations represented by $\mathbf{X}_t$ and $\mathbf{Z}_t$ are replicated f_t times. The essence matrix, $\ddot{\mathbf{Z}}$, is now

$$\ddot{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}_1 & & & \\ & \mathbf{Z}_2 & & \\ & & \ddots & \\ & & & \mathbf{Z}_T \end{bmatrix} \qquad (4.4.2)$$

where $\ddot{\mathbf{Z}}$ is a $N_e \times (cT)$ matrix. Note that the number of columns of $\mathbf{Z}$ do not equal the number of columns of $\ddot{\mathbf{Z}}$ as in Case II.

Returning to the Case III clinical trial example, suppose the trial of interest will last four years and observations will be taken yearly. Enrollment of subjects into the two treatment arms, standard treatment and new treatment, will occur during the first three years of the study only. Therefore, subjects entering within the first year of the study will have three years of evaluation. Those entering after the first year will have less than three years of follow-up. It will again be assumed that the dependent variable of interest varies linearly with time and a separate regression will be fit for each treatment. The fixed effect parameters are given by β_{01}, β_1, β_{02}, and β_2 where β_{0i} and β_i are the intercept and slope parameters for the standard treatment and the new treatment, respectively. The four times of data collection are denoted x_1, x_2, x_3, and x_4. Allowing each subject their own random effect parameter u_j, the essence matrices are

$$\ddot{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \\ \mathbf{X}_5 \\ \mathbf{X}_6 \end{bmatrix} = \left[\begin{array}{cccc} 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ 1 & x_3 & 0 & 0 \\ 1 & x_4 & 0 & 0 \\ \hline 0 & 0 & 1 & x_1 \\ 0 & 0 & 1 & x_2 \\ 0 & 0 & 1 & x_3 \\ 0 & 0 & 1 & x_4 \\ \hline 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ 1 & x_3 & 0 & 0 \\ \hline 0 & 0 & 1 & x_1 \\ 0 & 0 & 1 & x_2 \\ 0 & 0 & 1 & x_3 \\ \hline 1 & x_1 & 0 & 0 \\ 1 & x_2 & 0 & 0 \\ \hline 0 & 0 & 1 & x_1 \\ 0 & 0 & 1 & x_2 \end{array}\right]$$

and

$$\ddot{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}_1 & & & & & \\ & \mathbf{Z}_2 & & & & \\ & & \mathbf{Z}_3 & & & \\ & & & \mathbf{Z}_4 & & \\ & & & & \mathbf{Z}_5 & \\ & & & & & \mathbf{Z}_6 \end{bmatrix} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

In Case II, $\mathbf{R} = \mathbf{I}(f) \otimes \mathbf{R}^*$ where $\mathbf{R}^*$ is a $r \times r$ matrix. Now that the r_i vary between subjects in Case III, there is not a general $\mathbf{R}^*$. It will now be assumed that $\mathbf{R}$ is

$$\mathbf{R} = \text{Diag}\left(\mathbf{I}(f_1) \otimes \mathbf{R}_1^*, \mathbf{I}(f_2) \otimes \mathbf{R}_2^*, \cdots \mathbf{I}(f_T) \otimes \mathbf{R}_T^*\right). \tag{4.4.3}$$

The Diag operator diagonalizes the matrices $\mathbf{I}(f_t) \otimes \mathbf{R}_t^*$ for t=1...T. Let $\mathbf{R}_M^*$ be the $M \times M$ matrix for the maximum r_t. For r_t less than the maximum r_t, $\mathbf{R}_t^*$ is formed by using the appropriate rows and columns of $\mathbf{R}_M^*$. Additionally, $\mathbf{R}_i^* = \mathbf{R}_j^*$ when $r_i = r_j$. The essence matrix, $\ddot{\mathbf{R}}$, is

$$\ddot{\mathbf{R}} = \text{Diag}\left(\mathbf{R}_1^*, \mathbf{R}_2^*, \cdots \mathbf{R}_T^*\right). \tag{4.4.4}$$

$\mathbf{G}$ and $\ddot{\mathbf{G}}$ remain unchanged from Case II, equation 4.3.7 and 4.3.8, respectively. The weight matrix, $\mathbf{W}$, now has the form

$$\mathbf{W} = \begin{bmatrix} \frac{f_1}{N}\mathbf{I}(r_1) & & & \\ & \frac{f_2}{N}\mathbf{I}(r_2) & & \\ & & \ddots & \\ & & & \frac{f_T}{N}\mathbf{I}(r_T) \end{bmatrix}. \tag{4.4.5}$$

Thus for Case III, it can be shown that the inverted portion of $\mathbf{C}_{11}$ can be written in its distinct components as

$$\mathbf{X}'(\mathbf{ZGZ}' + \mathbf{R})^{-1}\mathbf{X} = N\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}' + \ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}} \tag{4.4.6}$$

The proof is the similar to the one given in Case II but the $\mathbf{Z}^*$ are now replaced by their corresponding $\mathbf{Z}_t$'s and the $\mathbf{R}^*$'s are now replaced by their corresponding $\mathbf{R}_t^*$'s.

Proof of Equality 4.4.6

First, working with the inner piece of $\mathbf{X}'(\mathbf{ZGZ}'+\mathbf{R})^{-1}\mathbf{X}$ leads to

$$
\begin{aligned}
&(\mathbf{ZGZ}'+\mathbf{R}) \\
&= \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{Z}_T \end{bmatrix} \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{G}^* & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{G}^* \end{bmatrix} \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{Z}_1' & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{Z}_T' \end{bmatrix} \\
&\quad + \begin{bmatrix} \mathbf{I}(f_1)\otimes\mathbf{R}_1^* & & \\ & \ddots & \\ & & \mathbf{I}(f_T)\otimes\mathbf{R}_T^* \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_T \end{bmatrix} \quad \text{where } \mathbf{A}_i = \left(\mathbf{I}(f_i)\otimes\mathbf{Z}_i\right)\left(\mathbf{I}(f_i)\otimes\mathbf{G}^*\right)\left(\mathbf{I}(f_i)\otimes\mathbf{Z}_i'\right)+\left(\mathbf{I}(f_i)\otimes\mathbf{R}_i^*\right) \\
&= \begin{bmatrix} \left(\mathbf{I}(f_1)\otimes\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}_1'\right)+\left(\mathbf{I}(f_1)\otimes\mathbf{R}_1^*\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}_T'\right)+\left(\mathbf{I}(f_T)\otimes\mathbf{R}_T^*\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}_1'+\mathbf{R}_1^*\right)\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}_T'+\mathbf{R}_T^*\right)\right) \end{bmatrix}
\end{aligned}
$$

This piece inverted is then given by,

$$
(\mathbf{ZGZ}'+\mathbf{R})^{-1} = \begin{bmatrix} \left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}_1'+\mathbf{R}_1^*\right)^{-1}\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}_T'+\mathbf{R}_T^*\right)^{-1}\right) \end{bmatrix}.
$$

So that,

$$\mathbf{X}'(\mathbf{ZGZ}'+\mathbf{R})^{-1}\mathbf{X}$$

$$=\begin{bmatrix}\mathbf{1}'_{f_1}\otimes\mathbf{X}'_1 & \cdots & \mathbf{1}'_{f_T}\otimes\mathbf{X}'_T\end{bmatrix}\begin{bmatrix}\left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1}\right) & & \\ & \ddots & \\ & & \left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\right)\end{bmatrix}$$

$$\begin{bmatrix}\mathbf{1}_{f_1}\otimes\mathbf{X}_1\\ \vdots\\ \mathbf{1}_{f_T}\otimes\mathbf{X}_T\end{bmatrix}$$

$$=\left(\mathbf{1}'_{f_1}\otimes\mathbf{X}'_1\right)\left(\mathbf{I}(f_1)\otimes\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1}\right)\left(\mathbf{1}_{f_1}\otimes\mathbf{X}_1\right)+\cdots$$

$$+\left(\mathbf{1}'_{f_T}\otimes\mathbf{X}'_T\right)\left(\mathbf{I}(f_T)\otimes\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\right)\left(\mathbf{1}_{f_T}\otimes\mathbf{X}_T\right)$$

$$=\left(\mathbf{1}'_{f_1}\mathbf{I}(f_1)\mathbf{1}_{f_1}\otimes\mathbf{X}'_1\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1}\mathbf{X}_1\right)+\cdots+\left(\mathbf{1}'_{f_T}\mathbf{I}(f_T)\mathbf{1}_{f_T}\otimes\mathbf{X}'_T\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\mathbf{X}_T\right)$$

$$=\left(f_1\otimes\mathbf{X}'_1\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1}\mathbf{X}_1\right)+\cdots+\left(f_T\otimes\mathbf{X}'_T\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\mathbf{X}_T\right)$$

$$=f_1\mathbf{X}'_1\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1}\mathbf{X}_1+\cdots+f_T\mathbf{X}'_T\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\mathbf{X}_T$$

$$=N\mathbf{X}'_1\frac{f_1}{N}\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1}\mathbf{X}_1+\cdots+N\mathbf{X}'_T\frac{f_T}{N}\left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\mathbf{X}_T$$

$$=N\begin{bmatrix}\mathbf{X}'_1 & \cdots & \mathbf{X}'_T\end{bmatrix}\begin{bmatrix}\frac{f_1}{N}\mathbf{I}(r_1) & & \\ & \ddots & \\ & & \frac{f_T}{N}\mathbf{I}(r_T)\end{bmatrix}\begin{bmatrix}\left(\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1\right)^{-1} & & \\ & \ddots & \\ & & \left(\mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_T\end{bmatrix}$$

$$=N\begin{bmatrix}\mathbf{X}'_1 & \cdots & \mathbf{X}'_T\end{bmatrix}\begin{bmatrix}\frac{f_1}{N}\mathbf{I}(r_1) & & \\ & \ddots & \\ & & \frac{f_T}{N}\mathbf{I}(r_T)\end{bmatrix}\begin{bmatrix}\mathbf{Z}_1\mathbf{G}^*\mathbf{Z}'_1+\mathbf{R}^*_1 & & \\ & \ddots & \\ & & \mathbf{Z}_T\mathbf{G}^*\mathbf{Z}'_T+\mathbf{R}^*_T\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_1\\ \vdots\\ \mathbf{X}_T\end{bmatrix}$$

$$= N\left[\mathbf{X}_1' \quad \cdots \quad \mathbf{X}_T'\right]\begin{bmatrix} \frac{f_1}{N}\mathbf{I}(r_1) & & \\ & \ddots & \\ & & \frac{f_T}{N}\mathbf{I}(r_T) \end{bmatrix}$$

$$\left\{\begin{bmatrix} \mathbf{Z}_1 & & \\ & \ddots & \\ & & \mathbf{Z}_T \end{bmatrix}\begin{bmatrix} \mathbf{G}^* & & \\ & \ddots & \\ & & \mathbf{G}^* \end{bmatrix}\begin{bmatrix} \mathbf{Z}_1' & & \\ & \ddots & \\ & & \mathbf{Z}_T' \end{bmatrix} + \begin{bmatrix} \mathbf{R}_1^* & & \\ & \ddots & \\ & & \mathbf{R}_T^* \end{bmatrix}\right\}^{-1}\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_T \end{bmatrix}$$

$$= N\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}' + \ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}$$

End of Proof for 4.4.6

Therefore, the noncentrality parameter for Case III can be written as $\lambda = N\lambda_{III}^*$ where λ_{III}^* is not related to N. Analogous to Case II, N stands for the total number of observed responses.

4.5 Summary of Noncentrality Findings

As stated in 4.1.3, the noncentrality parameter for the approximate noncentral F used for the testing of hypotheses regarding the fixed effects of the mixed linear model is

$$\lambda = \boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\mathbf{X}'(\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})^{-1}\mathbf{X}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}.$$

In Section 4.2, it was shown that the noncentrality parameter for a study design that follows Case I can be rewritten as

$$\lambda = N\boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\ddot{\mathbf{X}}'\mathbf{W}\left(N\ddot{\mathbf{Z}}\mathbf{G}\ddot{\mathbf{Z}}'\mathbf{W} + \ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}$$

$$= N\lambda_1^*.$$

In Sections 4.3 and 4.4, it was shown that the noncentrality parameter for a study design that follows Case II or Case III can be rewritten as

$$\lambda = N\boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}' + \ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}$$

$$= N\lambda_{II}^{*}$$

or

$$= N\lambda_{III}^{*}.$$

Therefore, a primary noncentrality parameter can be defined for each of the three cases discussed. However, the primary noncentrality does not depend on N, the total number of observed responses, for only Cases II and III. For these cases, the primary noncentrality is based solely on the design points to be used ($\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$), the weights for those points ($\mathbf{W}$), and the conjectured values for $\boldsymbol{\beta}$, $\ddot{\mathbf{G}}$, and $\ddot{\mathbf{R}}$. These primary noncentrality parameters only need to be calculated once for determining power with different sample sizes. Even though the primary noncentrality parameter for Case I is not independent of N, there is still a benefit to calculating the primary noncentrality parameter. The primary noncentrality for Case I is a more computationally efficient version of λ since the matrices being inverted have smaller dimension and all pieces of λ_I^* are fixed except for N.

The following chapter includes examples of calculating power using the results from this chapter. A simulation study based on mixed linear model power is presented in Chapter 6.

Chapter 5
Applications of Mixed Linear Model Power

5.1 Introduction

In this chapter, applications of the methods derived in Chapter 4 for calculating power for the mixed linear model are illustrated. Based on the approximate F, the power of a size α test of the fixed effects is given by $\Pi=P[F(\text{rank}(\mathbf{K}), N\text{-rank}(\mathbf{X}\ \mathbf{Z}), \lambda)\geq F_{\alpha}]$. The noncentrality parameter, λ, needed for calculating power can be written in its distinct components yielding computationally efficient versions of λ. The form of the primary noncentrality parameter varies depending on the design of the study. Thus, three cases are considered and their primary noncentrality parameters are found to be

Case I

$$\lambda_{\mathrm{I}}^{*}=\boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\ddot{\mathbf{X}}'\mathbf{W}\left(N\ddot{\mathbf{Z}}\mathbf{G}\ddot{\mathbf{Z}}'\mathbf{W}+\ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}$$

Case II or Case III

$$\lambda_{\mathrm{II}}^{*}=\lambda_{\mathrm{III}}^{*}=\boldsymbol{\beta}'\mathbf{K}\left\{\mathbf{K}'\left[\ddot{\mathbf{X}}'\mathbf{W}\left(\ddot{\mathbf{Z}}\ddot{\mathbf{G}}\ddot{\mathbf{Z}}'+\ddot{\mathbf{R}}\right)^{-1}\ddot{\mathbf{X}}\right]^{-1}\mathbf{K}\right\}^{-1}\mathbf{K}'\boldsymbol{\beta}.$$

The remaining sections of this chapter contain examples for each of the three cases. The parameter estimates of $\boldsymbol{\beta}$, $\mathbf{G}$, and $\mathbf{R}$ for most of the examples are based on hypothetical situations. It is noted when real data are used for the parameter estimates, however, the situations of interest are still hypothetical. The programs used to calculate the

primary noncentrality parameter and then powers for each of the examples can be found in Appendix A.

5.2 Case I Examples

Case I consists of situations where the column space for the random effects or the $\mathbf{Z}$ matrix remains fixed upon the addition of another subject. In order to calculate power at various sample sizes, one needs to specify the unique design points, $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$; the weighting of these points, $\mathbf{W}$; the estimates for $\boldsymbol{\beta}$, $\mathbf{G}$, and $\ddot{\mathbf{R}}$; the contrast matrix of interest, $\mathbf{K}$; and the type I error, α.

Example 5.2.1

Let's return to the three center clinical trial example first described in Section 4.2. In this example, two fixed treatments are going to be studied and the random effects are center and the center by treatment interactions. Thus, the essence model matrices are

$$\ddot{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \ddot{\mathbf{Z}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The response of interest has treatment means of 50.95 and 51.96 for treatment 1 and treatment 2, respectively. Therefore, $\boldsymbol{\beta}$ has the form

$$\boldsymbol{\beta} = \begin{bmatrix} 50.95 \\ 51.96 \end{bmatrix}.$$

Based on data from a previous clinical trial that studied similar treatments and centers, it will be assumed that the overall error variance is σ^2=0.0485. The variance contribution of center is $\sigma_c^2 = 0.1073$ and the center by treatment interaction is $\sigma_{ct}^2 = 0.0510$. These values come from the estimates of center mean square and the interaction mean square from the previous study. Therefore,

$$\ddot{\mathbf{R}} = \sigma^2 \mathbf{I}(6)$$

and

$$\mathbf{G} = \mathbf{I}(6) \otimes \begin{bmatrix} \sigma_c^2 & 0 & 0 \\ 0 & \sigma_{ct}^2 & 0 \\ 0 & 0 & \sigma_{ct}^2 \end{bmatrix}.$$

In order to compare the two treatment means, the contrast is

$$\mathbf{K}' = [1 \quad -1].$$

Assuming equal weighting of the 6 cells, the weight matrix is

$$\mathbf{W} = \tfrac{1}{6}\mathbf{I}(6).$$

For the case of equal weightings, each time a subject is added to one treatment a subject is added to the other treatment. So the total sample size will be divisible by 6.

Figure 5.1 contains the power curves for α=0.05 and 0.01. From Figure 5.1, 12 total subjects need to be sampled to achieve a power of at least 0.80 when α=0.05. This corresponds to 2 subjects for each treatment at each center and has an actual power of 0.961. If α=0.01, 18 total subjects are needed to achieve a power of at least 0.80. This

corresponds to sampling 3 subjects for each treatment at each center. The actual power for this situation is 0.934.

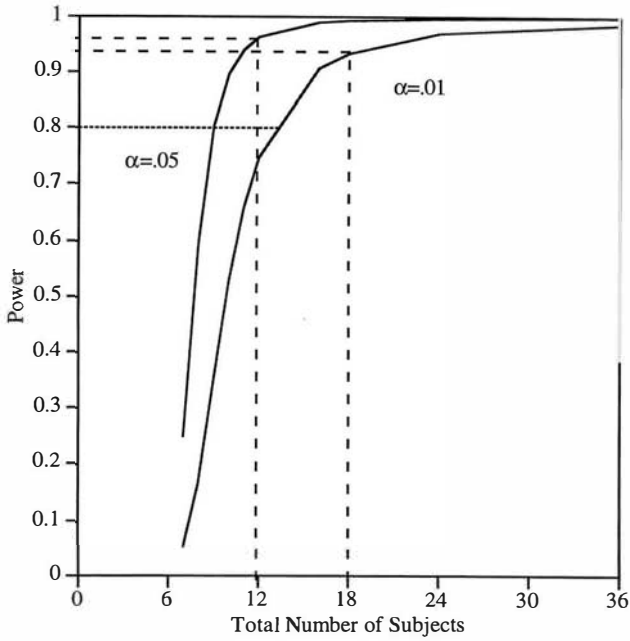

Figure 5.1: Plot of Power Curves for Case I Example 1

Example 5.2.2

A toxicologist would like to study the purity of a substance used in his lab experiments. This substance is purchased from three different suppliers. From each supplier, he will select four random batches and within each batch he will randomly select samples. Therefore, the random effects are the nested batches within each of the suppliers

and are given by u_{11}, u_{21}, u_{31}, u_{12}, u_{22}, u_{32}, u_{13}, u_{23}, u_{33}, u_{14}, u_{24}, and u_{34} where u_{ij} is the random effect for the j^{th} batch from the i^{th} supplier. The essence matrices are

$$\ddot{\mathbf{X}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \ddot{\mathbf{Z}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The toxicologist is willing to assume, from prior knowledge, purities of 10.31, 16.09, and 19.37 from Suppliers 1, 2, and 3, respectively. Therefore,

$$\boldsymbol{\beta} = \begin{bmatrix} 10.31 \\ 16.09 \\ 19.37 \end{bmatrix}.$$

He wants to perform a balanced and complete experiment, so $\mathbf{W} = \left(\frac{1}{12}\right)\mathbf{I}(12)$. For the $\mathbf{R}$ matrix, a simple variance structure of $\ddot{\mathbf{R}} = 3.3 \cdot \mathbf{I}(12)$ will be assumed. The toxicologist feels that the variance among batches varies from supplier to supplier and assumes

$$\mathbf{G}_1 = \mathbf{I}(4) \otimes \begin{bmatrix} 0.42 & 0 & 0 \\ 0 & 26.7 & 0 \\ 0 & 0 & 56.94 \end{bmatrix}.$$

The contrast matrix,

$$\mathbf{K}' = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix},$$

will be used to compare the purity of the substance from the three suppliers. He would like to see what kind of power he will have under this situation. On a whim, he would also like to see what the power would be if he assumed incorrectly homogeneity of the variances. In this case, the matrix $\mathbf{G}_2 = 26.8 \cdot \mathbf{I}(12)$ is assumed. The following plot contains the power curves under each of these situations.

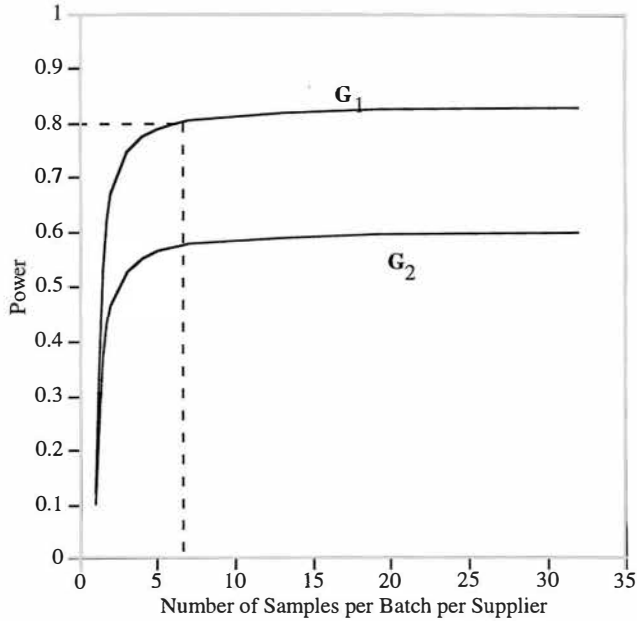

Figure 5.2: Plot of Power Curves for Case I Example 2

From Figure 5.2, the toxicologist can see that he will need to take 7 samples of each batch from each supplier to have a power of 0.805 when he assumes the heteroscedastic variances ($\mathbf{G}_1$). If the homoscedastic variance is assumed ($\mathbf{G}_2$), he would only have a power of 0.578 if he took 7 samples of each batch from each supplier. Even if the toxicologist took 40 samples of each batch from each supplier, he would only have a power of 0.600.

This example brings about an interesting question. Why do the two power curves asymptote to values other than one? Recall that for Case I, λ_I^* depends on **N**. So when λ_I^* is multiplied by N to get λ, a portion of the effect of N is absorbed by the N that is contained in λ_I^*. Thus, λ_I^* asymptotes. Upon further investigation, it is seen that for $\mathbf{G}_1$, $\underset{N\to\infty}{\lambda} = 10.5397$. Also as $N \to \infty$, F_α approaches a value near 2.9975. Thus, as $N \to \infty$, F_α=2.9975, λ=10.5397, rank(K)=2, and N-rank[**X Z**]$\to\infty$, power approaches a value of 0.836. Similarly, it is seen for $\mathbf{G}_2$ that $\underset{N\to\infty}{\lambda} = 6.28112$. So for $\mathbf{G}_2$, power approaches a value of 0.605.

5.3 Case II Examples

Case II consists of situations where the column space for the random effects increases with the addition of another subject. In order to calculate power for various sample sizes, one needs to specify the unique design points which are given by $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$; the weighting of these points, **W**; the estimates for $\boldsymbol{\beta}$, $\ddot{\mathbf{G}}$, and $\ddot{\mathbf{R}}$; the contrast matrix of interest, **K**; and the type I error, α.

Example 5.3.1

Suppose the longitudinal study described in Section 4.3 is going to be carried out by an investigator. The essence model matrices are

$$\ddot{\mathbf{X}} = \left[\begin{array}{cccc} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right] \quad \text{and} \quad \ddot{\mathbf{Z}} = \left[\begin{array}{c:c} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hdashline 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{array}\right].$$

Assuming that the dependent variable varies linearly with time, a separate regression will be fit for each treatment. From a pilot study, the estimates for the fixed effects are estimated as 21.2, 1.4, 20.3, and .95 which correspond to the intercept and slope for treatment 1 and treatment 2, respectively. Thus,

$$\boldsymbol{\beta} = \begin{bmatrix} 21.2 \\ 1.4 \\ 20.3 \\ .95 \end{bmatrix}.$$

It is assumed that there is a homogeneous variance structure for each subject so that

$$\ddot{\mathbf{G}} = \begin{bmatrix} 3.0 & 0 \\ 0 & 3.0 \end{bmatrix}.$$

Again from the pilot study, the investigator found that an unstructured covariance best described each subject's block contribution to the **R** matrix. Therefore using those estimates he will assume,

$$\ddot{\mathbf{R}} = \mathbf{I}(2) \otimes \begin{bmatrix} 2.4 & -0.2 & 0.9 \\ -0.2 & 1.2 & 0.01 \\ 0.9 & 0.01 & 3.5 \end{bmatrix}.$$

The investigator does not think he will have trouble enrolling subjects into his clinical trial, so he will assume equal sample sizes for each treatment and

$$\mathbf{W} = \tfrac{1}{6}\mathbf{I}(6).$$

The focus of his research is to compare both the intercepts and the slopes of the two treatments. Thus,

$$\mathbf{K}' = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Therefore, the primary noncentrality has a value of .0320459. Figure 5.3 contains the power curves for α=0.05 and 0.01.

From the plot in Figure 5.3, a total of 104 subjects or 52 subjects per treatment are needed to achieve an actual power of 0.811 when α=0.05. At α=0.01, one would have to sample 74 subjects per treatment to achieve a power of 0.807 at α=0.01. The investigator is quite pleased with these findings. He doesn't think he will have any problem with sampling the 52 subjects per treatment. His grant has also been renewed so he may even try to sample more subjects per treatment so that his power will be closer to 0.90.

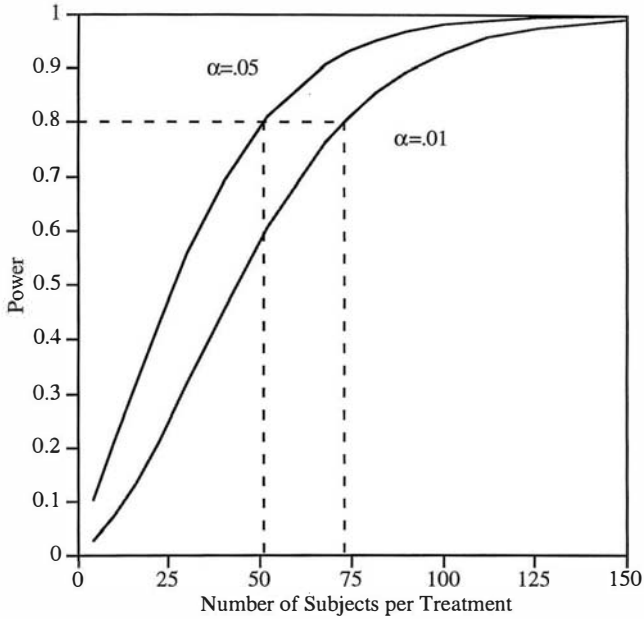

Figure 5.3: Plot of Power Curves for Case II Example 1

Example 5.3.2

Suppose a study comparing two blood pressure cuffs is planned. Three repeated measurements with each of two cuffs (new and standard) will be taken on each subject. Blood pressure "cuff" type is the fixed effect and the parameters are designated as β_{New} and β_{Std}. The random effect is subject×cuff so each subject has random effect parameters u_{iN} and u_{iS}. The essence model matrices are

$$\ddot{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } \ddot{\mathbf{Z}} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

The error matrix, **R**, will be modelled to allow for a different variance for each cuff:

$$\ddot{\mathbf{R}} = \begin{bmatrix} \sigma^2_{New} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{New} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{New} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{Std} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{Std} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{Std} \end{bmatrix}.$$

The random effects error matrix, **G**, will be modelled with an unstructured covariance matrix:

$$\ddot{\mathbf{G}} = \begin{bmatrix} \sigma^2_1 & \sigma_{12} \\ \sigma_{12} & \sigma^2_2 \end{bmatrix}.$$

Medical investigators from the School of Nursing at the Medical College of Virginia performed a similar study in which they compared the Johnson blood pressure cuff to the standard blood pressure cuff. The new experiment will be using a "modified" Johnson cuff. It is anticipated that the parameter estimates from the previous study will be close to the data from the new study except for an increase of 2% in the new cuff mean from the Johnson blood pressure cuff mean. Therefore, the estimates that will be used are

β_{New}=121.06×1.02=123.5, β_{Std}=120.47,

σ^2_{New}=42.36, σ^2_{Std}=42.31,

σ_1^2=325.33, σ_2^2=321.32, and σ_{12}=328.58.

Since each subject will have three measurements from each of the blood pressure cuffs, the weight matrix is $\mathbf{W} = \frac{1}{6}\mathbf{I}(6)$. In order to compare the means of the two cuffs, $\mathbf{K}' = [1 \quad -1]$. The primary noncentrality based on these matrices has a value of 0.0863841. The following plot shows the power curves for α=0.05 and 0.01.

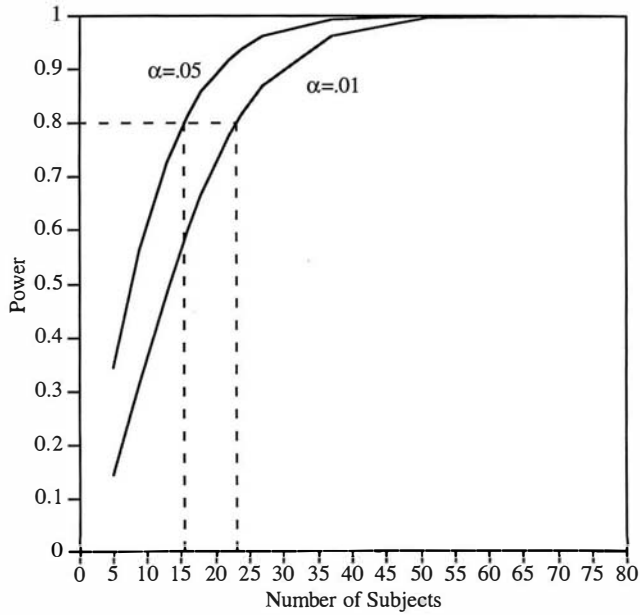

Figure 5.4: Plot of Power Curves for Case II Example 2

From Figure 5.4 , one can see that 16 subjects need to be sampled at α=0.05 and 24 subjects need to be sampled at α=0.01 in order to achieve a power of at least 0.80. These values correspond to approximate powers of 0.81333 and 0.81854, respectively.

5.4 Case III Examples

Case III consists of situations where the column space for the random effects increases with the addition of another subject as in Case II. Unlike Case II, Case III will allow for the row space for the fixed and random effects to vary from subject to subject. As in Case II, one needs to specify the unique design points which are given by $\ddot{\mathbf{X}}$ and $\ddot{\mathbf{Z}}$; the weighting of these points, **W**; the estimates for $\boldsymbol{\beta}$, $\ddot{\mathbf{G}}$, and $\ddot{\mathbf{R}}$; the contrast matrix of interest, **K**; and the type I error, α to calculate power for various sample size.

Example 5.4.1

An investigator wants to perform the clinical trial discussed in Section 4.4. Enrollment of subjects into either the standard treatment or the new treatment will occur during the first three years of the four years planned for the study. Observations will be taken once each year from enrollment to study completion. It is assumed that the dependent variable varies linearly with time and a separate regression will be fit for each treatment. With subject as the random effect, the essence model matrices are

$$\ddot{\mathbf{X}} = \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 4 \\ \hline 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 3 \\ \hline 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad \text{and} \quad \ddot{\mathbf{Z}} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

The estimates for the intercept and slope for each treatment are assumed to be -5, 55.5, -4.5, and 52.3, respectively. Thus,

$$\boldsymbol{\beta} = \begin{bmatrix} -5 \\ 55.5 \\ -4.5 \\ 52.3 \end{bmatrix}.$$

Homogeneity between subjects is also assumed so that

$$\ddot{\mathbf{G}} = 0.39 \cdot \mathbf{I}(6).$$

Assuming an autoregressive structure for the blocks of the **R** matrix and using data from a previous study, the largest block (corresponding to the block for a subject with all four measurements) is

$$\mathbf{R}_4^* = \begin{bmatrix} 1905 & 20.8 & 0.23 & 0.002 \\ 20.8 & 1905 & 20.8 & 0.23 \\ 0.23 & 20.8 & 1905 & 20.8 \\ 0.002 & 0.23 & 20.8 & 1905 \end{bmatrix}.$$

The block for a subject with three years of observations is denoted by $\mathbf{R}_3^*$ and is made from the first three rows and three columns of $\mathbf{R}_4^*$. In a similar fashion, the block for a subject with two years of observations is denoted by $\mathbf{R}_2^*$. Using $\mathbf{R}_4^*$, $\mathbf{R}_3^*$, and $\mathbf{R}_2^*$, the form of $\ddot{\mathbf{R}}$ is

$$\ddot{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_4^* & & & & & \\ & \mathbf{R}_4^* & & & & \\ & & \mathbf{R}_3^* & & & \\ & & & \mathbf{R}_3^* & & \\ & & & & \mathbf{R}_2^* & \\ & & & & & \mathbf{R}_2^* \end{bmatrix}.$$

The contrast matrix is constructed to compare both the intercepts and slopes for each treatment where the contrast matrix $\mathbf{K}$ is

$$\mathbf{K}' = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

Two sampling schemes have been suggested for this clinical trial. First, suppose that 50% of the subjects are enrolled during the first year, $33\frac{1}{3}\%$ of the subjects are enrolled during the second year, and $16\frac{2}{3}\%$ of the subjects are enrolled during the third year. Second, an alternative case of equal recruitment for each of the 4 study years has also been suggested. The weight matrices for each of these cases are

$$\mathbf{W}_1 = \tfrac{1}{18}\mathbf{I}(18)$$

for equal weighting of subjects across each of the three years that enrollment is allowed, and

$$\mathbf{W}_2 = \begin{bmatrix} \frac{3}{40}\mathbf{I}(8) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{2}{40}\mathbf{I}(6) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{40}\mathbf{I}(4) \end{bmatrix}$$

for the unequal weighting plan. The primary noncentrality has a value of 0.0063749 when using $\mathbf{W}_1$ and has a value of 0.007249 when using $\mathbf{W}_2$. The following plot contains the power curves for each of the weight matrices.

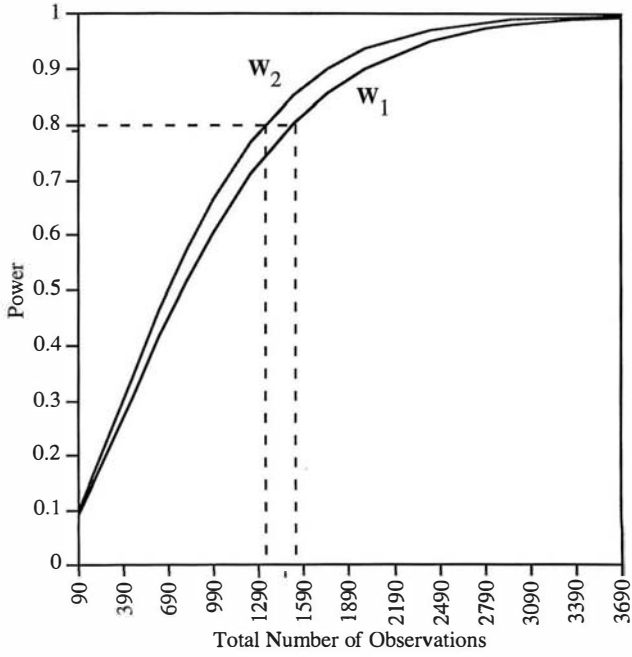

Figure 5.5: Plot of Power Curves for Case III Example 1

The horizontal axis in Figure 5.5 corresponds to the total number of observations necessary to achieve the corresponding power. In order to determine the number of subjects needed each year for each treatment, one needs to multiply this value by the weight matrix. Thus, to achieve a power of 0.80 when one assumes an equal number of subjects enrolling per year, a total of 1530 observations are necessary. This corresponds to 85 subjects per year per treatment (derived from $1530 \times \mathbf{W}_1$). When one wants 50% of the subjects from year 1, $33\frac{1}{3}$% of the subjects from year 2, and $16\frac{2}{3}$% of the subjects from year 3, a total of 1360 observations are necessary to achieve a power of 0.80. This corresponds to 102 subjects per treatment in year 1, 68 subjects per treatment in year 2, and 34 subjects per treatment in year 3 (derived from $1360 \times \mathbf{W}_2$).

Example 5.4.2

Suppose one is interested in studying a drug at two dose levels and a placebo. The objective will be to show that the doses of the drug have a different outcome trend than the placebo. The study will run for five continuous weeks. In order to save on the cost of the treatment, and hopefully, cut back on the number of subject missing a week, some subjects will be evaluated each week and some subjects will be evaluated only during the first, third and fifth weeks of the study. It will be assumed that the dependent variable of interest varies linearly with time and a separate regression will be fit for each dose level. The essence model matrices are

$$\ddot{\mathbf{X}} = \left[\begin{array}{cccccc}
1 & -2 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 1 & -2 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 2 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 2 \\
\hline
1 & -2 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 1 & -2 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 2 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 2
\end{array}\right]$$

and

$$\ddot{\mathbf{Z}} = \left[\begin{array}{cc|cc|cc|cc|cc|cc}
1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2
\end{array}\right].$$

The parameter estimates being used for the power analysis are based on the investigator's previous experience with the highest dose level of the drug. The estimates for the low dose level have been linearly interpolated from the placebo and high dose level values. Therefore, the intercept and slope estimates are assumed to be -1.39 and -0.035 for placebo, -2.39 and -0.176 for Dose 1, and -3.38 and -0.318 for Dose 2, so that

$$\boldsymbol{\beta} = \left[-1.39 \quad -0.025 \quad -2.39 \quad -0.176 \quad -3.38 \quad -0.318\right]'.$$

An unstructured covariance is assumed for each subject's random effects or

$$\ddot{\mathbf{G}} = \mathbf{I}(6) \otimes \begin{bmatrix} 1.15 & 0.163 \\ 0.163 & 0.039 \end{bmatrix},$$

and a simple covariance is assumed for the blocks of $\mathbf{R}$ or

$$\ddot{\mathbf{R}} = \begin{bmatrix} 0.125 \cdot \mathbf{I}(15) & \\ & 0.125 \cdot \mathbf{I}(9) \end{bmatrix}.$$

The contrast to compare the slope of the placebo to the slope of each dose level is

$$\mathbf{K}' = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Powers when an equal number of subjects are sampled for each of the two evaluation schemes and when 25% of the subjects are evaluated all five times and 75% are evaluated only during the first, third, and fifth weeks will be calculated. The weight matrices for these cases are

$$\mathbf{W}_1 = \tfrac{1}{24}\mathbf{I}(24)$$

and

$$\mathbf{W}_2 = \begin{bmatrix} \tfrac{1}{42}\mathbf{I}(15) & \mathbf{0} \\ \mathbf{0} & \tfrac{3}{42}\mathbf{I}(9) \end{bmatrix},$$

respectively. The primary noncentrality has a value of 0.0629176 when using $\mathbf{W}_1$ and has a value of 0.0708338 when using $\mathbf{W}_2$. The following plot contains the power curves for each of the weight matrices.

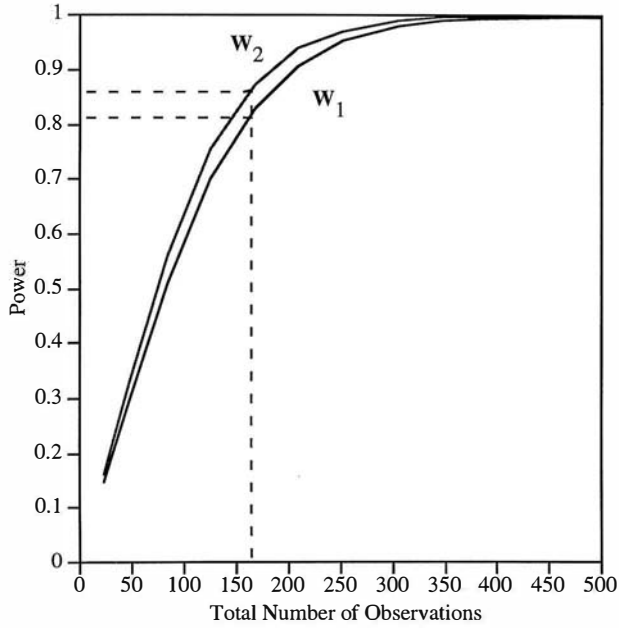

Figure 5.6: Plot of Power Curves for Case III Example 2

From Figure 5.6, to achieve a power of at least 0.80, a total of 168 observations are necessary when one assumes an equal number of subjects for each evaluation scheme. This corresponds to 7 subjects per evaluation scheme per dose level and has an actual power of 0.830. When one wants 25% of the subjects to be evaluated 5 times and 75% of the subjects to be evaluated only 3 times, a total of 168 observations are needed. This corresponds to 4 subjects per dose evaluated all 5 weeks and 12 subjects per dose evaluated during weeks 1, 3, and 5 and has an actual power of 0.873.

Chapter 6
Simulation Study

6.1 Introduction

A simulation study is conducted to accomplish two objectives. The first is to determine the effect on power when misspecification of the model's covariance structure occurs. As an example of misspecification, suppose the power analysis assumes a compound symmetry covariance when the sampled data actually follows an autoregressive covariance structure. For this example, the same number of parameters are assumed in the covariance matrix (2 parameters), however, the structure is incorrect. The second objective is to investigate whether sufficient power is achieved when the power analysis assumes a covariance structure with fewer parameters than the true covariance of the sampled data. This will be referred to as underspecification of the covariance. As an example of underspecification, suppose the power analysis assumes a compound symmetry covariance (2 parameters) and the true covariance of the sampled data has an unstructured covariance ($p(p+1)/2$ parameters).

A model based on a longitudinal study of two treatment groups is considered with this simulation. At each step of the simulation, the hypothesis is tested that the two treatment groups have differing trends over four measured occasions. A separate regression will be fit for each group, so the comparison of the two treatment group slopes is a test of the trends. This design follows a Case II situation discussed in Chapter 4. The program for the simulation study is written in SAS using Proc IML and Proc Mixed.

6.2 Data Generation

A primary focus of this research is the mixed linear model and its modelling with various covariance structures. As stated in Section 3.6, the mixed model is a useful tool for analyzing repeated measures or longitudinal data. For this simulation study, the **R** matrix will be used to model the covariance structure of a subject's data. Therefore, random normal data are generated with mean $\mathbf{X}\boldsymbol{\beta}$ and variance **R** where

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{1} \\ \beta_{02} \\ \beta_{2} \end{bmatrix}, \quad \text{and} \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}^{*} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{*} \end{bmatrix}$$

and β_{0i} and β_i are the intercept and slope parameters for a subject in the i^{th} group, and $\mathbf{R}^{*}$ is a 4×4 covariance matrix for a subject from either group.

To generate random normal data with mean $\mathbf{X}\boldsymbol{\beta}$ and variance **R**, first generate random numbers having a normal distribution with mean 0 and variance 1 using the NORMAL function in SAS. In order to generate data for n subjects per group, this is done with the following SAS statement,

```
z=NORMAL(REPEAT(seed, n, 8));
```

The result of this statement is a n×8 matrix whose elements are normally distributed with mean 0 and variance 1. Next, the SAS statement,

```
y=z*ROOT(R) + J(n,1,1)*(X*beta)`;
```

is used to obtain the random normal data with mean $\mathbf{X}\boldsymbol{\beta}$ and variance $\mathbf{R}$. The SAS function ROOT(matrix) performs the Cholesky decomposition of a symmetric, positive definite matrix. The resulting matrix, y, is n×8. Each row of this matrix corresponds to responses of one subject from group 1 and of one subject from group 2. The first four columns of this matrix correspond to the four responses of a subject from group 1 and the last four columns correspond to the four responses for a subject from group 2.

Several factors of interest are varied through the simulation. These factors are chosen because the nature of data analysis depends on which parameters have the most effect on power and are known with the least certainty. In turn, these values will help to address the objectives of the simulation. Therefore, the effects of varying 1) sample size, 2) the true difference, and 3) variance, the three components of the noncentrality parameter, are usually examined. The first factor of consideration for the simulation is sample size per group, n. Simulated values of n include n=10, n=20, and n=40. The second factor of interest is the difference between the slope parameters. The differences used are 0 (no difference), 0.2 (small difference), and 0.45 (medium/large difference). The first value of no difference is used for validation since when there is no difference in the slope parameters the power of the test should be equal to the alpha level which is 0.05. The small and medium/large difference values were chosen so that powers of interest could be achieved. The third factor considered is the variance structure assumed for $\mathbf{R}^*$. The three covariance structures used to generate the data include unstructured, compound symmetry, and autoregressive. The final factor varied in the simulation is the level of correlation. The levels of correlation considered are small (ρ=0.1), medium (ρ=0.5), and large (ρ=0.9). Only positive correlations are considered because for most repeated measurement situations, negative correlations are not expected (Chinchilli, 1996). Since the hypothesis of interest tests the slopes of the two groups, the values of the intercept parameters are not

very important, and thus are fixed at $\beta_{01} = 4.2$ and $\beta_{02} = 4.95$. In order to achieve the differences in the slopes described above, the slope values given in Table 6.1 are used.

Table 6.1

Values for Slope Parameters

Difference	β_1	β_2
No (0)	1.25	1.25
Small (0.2)	1.25	1.45
Med/Large (0.45)	1.25	1.7

The covariance structures used for $\mathbf{R}^*$ have the following forms:

Unstructured (UN)

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}$$

Compound Symmetry (CS)

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

and

Autoregressive (AR)

$$\sigma^2\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

For compound symmetry and autoregressive, a value of σ^2=1.5 is assumed for the variance term. For a given correlation, $\mathbf{R}^*$ is then formed using this value and the correlation value (ρ=0.1, 0.5, 0.9) assumed. For the unstructured covariance structure, values of $\sigma_{11} = 1$, $\sigma_{22} = 1.33$, $\sigma_{33} = 1.66$, and $\sigma_{44} = 2$ are assumed. In order to have positive definite matrices with the correlations varied in the simulation, the following matrices are used

UN ρ=0.1

$$\begin{bmatrix} 1 & .115 & .129 & .141 \\ .115 & 1.33 & .149 & .163 \\ .129 & .149 & 1.66 & .182 \\ .141 & .163 & .182 & 2 \end{bmatrix}$$

UN ρ=0.5

$$\begin{bmatrix} 1 & .577 & .644 & .707 \\ .577 & 1.33 & .743 & .815 \\ .644 & .743 & 1.66 & .911 \\ .707 & .815 & .911 & 2 \end{bmatrix}$$

UN ρ=0.9

$$\begin{bmatrix} 1 & .967 & 1.16 & 1.27 \\ .967 & 1.33 & 1.34 & 1.47 \\ 1.16 & 1.34 & 1.66 & 1.64 \\ 1.27 & 1.47 & 1.64 & 2 \end{bmatrix}.$$

With the three different values for n, three differences of the two slopes, three covariance structures assumed for $\mathbf{R}^*$, and three values of correlation, the resulting simulation study will consider 3×3×3×3=81 total cases. In each of these cases, 1000 repetitions are simulated and done as 10 runs of 100 sets. Due to time and computer restraints, it was decided that 1000 simulated data sets would be sufficient for drawing conclusions. In performing simulation studies, most statisticians do not go below 500 simulated data sets (Chinchilli, 1996). The 1000 data sets were done as 10 runs of 100 sets so that means and standard deviations could be calculated. For each of the 100 sets, the normal data with mean $\mathbf{X}\boldsymbol{\beta}$ and variance $\mathbf{R}$ is generated as described above. This data is then transposed into a univariate fashion so that SAS's Proc Mixed can be used for analysis. Next, the dataset is analyzed three times with Proc Mixed. Each run assumes one of the three covariance structures and the F statistic for the hypothesis of equal slopes is calculated. For each run, the number of times the hypothesis is rejected for each covariance structure is counted and the empirical power is then calculated by dividing this number by 100. Ten empirical powers for each of the three covariance structures result for each case. Thus, powers assuming the "correct" covariance structure and assuming two "incorrect" covariance structures are calculated. Finally, the mean, median, and confidence intervals about the mean are calculated for each of the ten empirical powers. Plots of the means are used to investigate the misspecification and underspecification of the covariance structure.

6.3 Simulation Analysis and Results

The results of the simulation are given in Tables 6.2 - 6.10. Each table provides the three different values of correlation (ρ) and the three different covariance structures assumed to generate the data. Tables 6.2 - 6.4 contain the results for n=10 and no difference, small difference, and medium/large difference, respectively. Tables 6.5 - 6.7

contain the results for n=20 and no difference, small difference, and medium/large difference, respectively. Tables 6.8 - 6.10 contain the results for n=40 and no difference, small difference, and medium/large difference, respectively. For each of the covariance structures, the tables include (1) the mean of the empirical power, (2) a 95% confidence interval about this mean, (3) the median of the empirical power, and (4) the minimum and maximum values of the empirical power. The bold values in these tables represent the results when the data was analyzed with the covariance structure under which the data was actually simulated. Plots of the mean values are contained in Figures 6.1-6.9. It should be noted that these plots are not all on the same scale, so interpretation across plots must be done carefully.

When there is no difference in the slope parameters, as n and ρ increase, powers approach the value of 0.05 in a decreasing fashion. See Tables 6.2, 6.5 and 6.8, and Figures 6.1, 6.4, and 6.7. When the data is generated with an autoregressive covariance structure and analyzed with an autoregressive covariance, power is closest to 0.05 for all sample sizes and correlations. However, when this data is analyzed with a compound symmetry covariance structure, the power inflates as the correlation increases. When assuming a compound symmetry covariance, the power is inflated by at least 0.01 at ρ=0.1 and increases to more than 0.1 at ρ=0.9 for all samples sizes. Powers calculated assuming an unstructured covariance seem to be slightly inflated even when the data is generated with an unstructured covariance. For a sample size of 10, powers found when analyzing the data with an unstructured covariance are around 0.1. At n=20, all powers are around 0.06 and it is not until n=40 and ρ=0.9 that the power is close to 0.05. Finally, when assuming an autoregressive covariance for data generated with an unstructured or compound symmetry covariance, power approaches zero as the correlation increases.

For small differences in the slope parameters (0.2), power increases as n and ρ increase as expected. See Tables 6.3, 6.6, and 6.9, and Figures 6.2, 6.5, and 6.8. When

n=20 and ρ=0.9, powers greater than 0.8 are achieved when analyzing unstructured or compound symmetry data with either of those covariance structures. Sufficient power for autoregressive data is not reached until n=40 and ρ=0.9. As n and ρ increase, powers for analyzing unstructured data with a compound symmetry covariance get closer to the powers found when analyzed with the true unstructured covariance. As in the case when there is no difference in the slope parameters, powers are inflated as ρ increases when analyzing autoregressive data with a compound symmetry covariance. The inflation of power ranges from 0.02 when ρ=0.1 to as much as 0.2 when ρ=0.9 across all sample sizes. When analyzed with the unstructured covariance, the powers are fairly close to those of the true autoregressive powers especially, as n and ρ increase. The difference in power when assuming the unstructured covariance for autoregressive data is at most 0.065, and this occurs with the smallest sample size per group. At lower levels of correlation (ρ=0.1 and 0.5), there is essentially no difference in powers between the three covariance structures when the data is generated with an autoregressive covariance structure. This difference does increase some when the data is generated with an unstructured covariance.

For a medium/large difference in the slope parameters (0.45), as n increases, the difference in power for the different levels of correlation decreases for all generated and analyzed covariance structures. See Tables 6.4, 6.7, and 6.10, and Figures 6.3, 6.6, and 6.9. When n=10, the maximum difference in power is 0.3 and this occurs when analyzing unstructured data with an autoregressive covariance at ρ=0.5. The maximum difference in power decreases to 0.15, and this is under the same situation except n=20. For n=40, powers greater than 0.95 are achieved with all levels of correlation. Even with a sample size of ten subjects per group, powers greater than 0.9 are achieved when analyzing data with the generated covariance structure. There is little difference between compound symmetry and unstructured powers even when the data is generated with a unstructured

covariance matrix. The largest difference in power is 0.075, and this occurs when n=10. The average difference in power is only .015. Analyzing autoregressive data with a compound symmetry structure still inflates power but this inflation of power is less as n increases.

Table 6.2

Simulation Results for n=10 and No Difference in Slope Parameters

Basis for Analysis		Simulation: UN	Simulation: CS	Simulation: AR
			ρ=0.1	
UN	(1)	**.100**	.103	.100
	(2)	**(.037, .162)**	(.036, .170)	(.034, .166)
	(3)	**.100**	.105	.09
	(4)	**.06, .14**	.02, .14	.07, .17
CS	(1)	.053	**.066**	.061
	(2)	(.007, .099)	**(.018, .114)**	(.008, .114)
	(3)	.05	**.06**	.06
	(4)	.03, .09	**.04, .11**	.03, .12
AR	(1)	.039	.044	**.048**
	(2)	(.005, .073)	(0.0, .091)	**(0.0, .102)**
	(3)	.04	.04	**.04**
	(4)	.01, .06	.01, .09	**.02, .11**
			ρ=0.5	
UN	(1)	**.100**	.105	.102
	(2)	**(.026, .174)**	(.038, .172)	(.060, .144)
	(3)	**.105**	.105	.095
	(4)	**.04, .14**	.05, .15	.07, .13
CS	(1)	.051	**.064**	.125
	(2)	(0.0, .105)	**(.021, .107)**	(.044, .206)
	(3)	.05	**.07**	.125
	(4)	.02, .09	**.02, .09**	.07, .21
AR	(1)	.007	.012	**.044**
	(2)	(0.0, .028)	(0.0, .030)	**(.019, .069)**
	(3)	0.0	.015	**.04**
	(4)	0.0, .03	0.0, .02	**.03, .07**
			ρ=0.9	
UN	(1)	**.095**	.102	.085
	(2)	**(.042, .148)**	(.052, .152)	(.022, .149)
	(3)	**.095**	.105	.09
	(4)	**.05, .13**	.06, .14	.03, .13
CS	(1)	.049	**.051**	.168
	(2)	(.008, .090)	**(.015, .087)**	(.078, .258)
	(3)	.05	**.055**	.17
	(4)	.02, .09	**.01, .07**	.07, .24
AR	(1)	.002	.002	**.054**
	(2)	(0.0, .010)	(0.0, .010)	**(0.0, .115)**
	(3)	0.0	0.0	**.055**
	(4)	0.0, .01	0.0, .01	**.01, .11**

Table 6.3

Simulation Results for n=10 and Small Difference in Slope Parameters

Basis for: Analysis		Simulation UN	CS	AR
			ρ=0.1	
UN	(1)	**.213**	.195	.188
	(2)	**(.158, .268)**	(.133, .257)	(.113, .263)
	(3)	**.21**	.19	.185
	(4)	**.17, .26**	.17, .27	.13, .24
CS	(1)	.151	**.140**	.142
	(2)	(.094, .208)	**(.049, .231)**	(.066, .218)
	(3)	.14	**.135**	.14
	(4)	.11, .20	**.06, .24**	.07, .21
AR	(1)	.127	.118	**.122**
	(2)	(.064, .190)	(.034, .2020	**(.057, .187)**
	(3)	.125	.12	**.125**
	(4)	.08, .18	.04, .19	**.08, .17**
			ρ=0.5	
UN	(1)	**.243**	.283	.190
	(2)	**(.173, .313)**	(.192, .374)	(.148, .232)
	(3)	**.240**	.265	.195
	(4)	**.18, .29**	.24, .39	.14, .21
CS	(1)	.205	**.204**	.244
	(2)	(.112, .298)	**(.097, .311)**	(.161, .3270
	(3)	.2	**.20**	.255
	(4)	.12, .27	**.11, .30**	.16, .31
AR	(1)	.074	.067	**.139**
	(2)	(.034, .114)	(.016, .118)	**(.089, .189)**
	(3)	.075	.06	**.14**
	(4)	0.0, .11	.02, .12	**.08, .17**
			ρ=0.9	
UN	(1)	**.609**	.745	.385
	(2)	**(.510, .708)**	(.671, .819)	(.305, .465)
	(3)	**.615**	.745	.39
	(4)	**.54, .68**	.67, .79	.33, .44
CS	(1)	.608	**.733**	.541
	(2)	(.561, .655)	**(.630, .836)**	(.467, .615)
	(3)	.605	**.74**	.525
	(4)	.57, .65	**.65, .83**	.50, .60
AR	(1)	.178	.193	**.325**
	(2)	(.128, .228)	(.120, .266)	**(.245, .405)**
	(3)	.17	.19	**.315**
	(4)	.14, .21	.15, .27	**.25, .39**

Table 6.4

Simulation Results for n=10 and Med/Large Difference in Slope Parameters

Basis for: Analysis		Simulation UN	CS	AR
		ρ=0.1		
UN	(1)	**.534**	.519	.482
	(2)	**(.444, .624)**	(.429, .609)	(.389, .575)
	(3)	**.525**	.535	.465
	(4)	**.49, .63**	.45, .58	.44, .60
CS	(1)	.472	**.444**	.480
	(2)	(.317, .627)	**(.361, .527)**	(.363, .597)
	(3)	.445	**.45**	.465
	(4)	.39, .65	**.35, .50**	.41, .61
AR	(1)	.429	.396	**.454**
	(2)	(.285, .573)	(.332, .460)	**(.340, .568)**
	(3)	.415	.395	**.435**
	(4)	.36, .60	.33, .45	**.39, .59**
		ρ=0.5		
UN	(1)	**.742**	.761	.524
	(2)	**(.638, .846)**	(.675, .847)	(.413, .635)
	(3)	**.750**	.755	.50
	(4)	**.65, .83**	.69, .82	.47, .65
CS	(1)	.704	**.730**	.613
	(2)	(.589, .819)	**(.606, .854)**	(.519, .707)
	(3)	.710	**.725**	.61
	(4)	.58, .78	**.63, .81**	.54, .70
AR	(1)	.443	.446	**.488**
	(2)	(.321, .656)	(.382, .510)	**(.397, .579)**
	(3)	.410	.435	**.47**
	(4)	.38, .54	.41, .51	**.43, .57**
		ρ=0.9		
UN	(1)	**.996**	1.00	.917
	(2)	**(.986, 1.00)**	(1.00, 1.00)	(.852, .982)
	(3)	**1.00**	1.00	.915
	(4)	**.99, 1.00**	1.00, 1.00	.87, .97
CS	(1)	.998	**1.00**	.972
	(2)	(.989, 1.00)	**(1.00, 1.00)**	(.933, 1.00)
	(3)	1.00	**1.00**	.98
	(4)	.99, 1.00	**1.00, 1.00**	.94, 1.00
AR	(1)	.929	.986	**.917**
	(2)	(.881, .977)	(.956, 1.00)	**(.850, .984)**
	(3)	.93	.99	**.925**
	(4)	.87, .96	.96, 1.00	**.85, .96**

Table 6.5

Simulation Results for n=20 and No Difference in Slope Parameters

Basis for:		Simulation		
		UN	CS	AR
Analysis			ρ=0.1	
UN	(1)	**.086**	.088	.063
	(2)	**(.028, 144)**	(.026, .150)	(.005, .121)
	(3)	**.08**	.09	.065
	(4)	**.05, .13**	.04, .14	.03, .11
CS	(1)	.054	**.065**	.061
	(2)	(.007, .100)	**(.017, .113)**	(0.0, .124)
	(3)	.05	**.065**	.065
	(4)	.02, .10	**.03, .10**	.02, .11
AR	(1)	.034	.047	**.050**
	(2)	(0.0, .069)	(.011, .082)	**(0.0, .112)**
	(3)	.035	.045	**.045**
	(4)	0.0, .06	.02, .07	**.01, .10**
			ρ=0.5	
UN	(1)	**.069**	.067	.074
	(2)	**(.026, .112)**	(.016, .118)	(.012, .136)
	(3)	**.07**	.065	.075
	(4)	**.04, .11**	.03, .12	.03, .12
CS	(1)	.060	**.045**	.121
	(2)	(.008, .112)	**(.009, .081)**	(.049, .193)
	(3)	.055	**.045**	.12
	(4)	.02, .12	**.02, .07**	.05, .18
AR	(1)	.001	.006	**.048**
	(2)	(0.0, .007)	(0.0, .025)	**(0.0, .103)**
	(3)	0.0	0.0	**.045**
	(4)	0.0, .01	0.0, .03	**.01, .09**
			ρ=0.9	
UN	(1)	**.063**	.065	.071
	(2)	**(.024, .102)**	(.029, .101)	(.023, .119)
	(3)	**.065**	.06	.07
	(4)	**.02, .09**	.04, .10	.04, .11
CS	(1)	.059	**.056**	.158
	(2)	(.029, .089)	**(.011, .101)**	(.085, .231)
	(3)	.055	**.055**	.15
	(4)	.04, .09	**.02, .08**	.12, .23
AR	(1)	.003	.001	**.042**
	(2)	(0.0, .012)	(0.0, .007)	**(0.0, .091)**
	(3)	0.0	0.0	**.05**
	(4)	0.0, .01	0.0, .01	**.01, .08**

Table 6.6

Simulation Results for n=20 and Small Difference in Slope Parameters

Basis for:		Simulation		
		UN	CS	AR
Analysis			ρ=0.1	
UN	(1)	**.264**	.234	.233
	(2)	**(.051, .164)**	(.129, .339)	(.148, .318)
	(3)	**.26**	.23	.245
	(4)	**.19, .35**	.13, .32	.15, .29
CS	(1)	.240	**.230**	.245
	(2)	(.121, .359)	**(.130, .330)**	(.164, .326)
	(3)	.24	**.225**	.25
	(4)	.13, .35	**.15, .31**	.17, .31
AR	(1)	.196	.199	**.213**
	(2)	(.093, .299)	(.100, .298)	**(.132, .294)**
	(3)	.19	.205,	**.21**
	(4)	.12, .31	.13, .28	**.15, .31**
			ρ=0.5	
UN	(1)	**.417**	.386	.249
	(2)	**(.336, .498)**	(.275, .497)	(.102, .396)
	(3)	**.42**	.385	.245
	(4)	**.35, .45**	.30, .45	.17, .43
CS	(1)	.386	**.363**	.337
	(2)	(.294, .478)	**(.289, .437)**	(.210, .464)
	(3)	.39	**.36**	.33
	(4)	.29, .45	**.28, .42**	.27, .48
AR	(1)	.147	.149	**.225**
	(2)	(.093, .201)	(.090, .208)	**(.108, .342)**
	(3)	.145	.15	**.21**
	(4)	.09, .18	.10, .19	**.15, .35**
			ρ=0.9	
UN	(1)	**.863**	.961	.548
	(2)	**(.822, .904)**	(.928, .994)	(.482, .615)
	(3)	**.855**	.965	.56
	(4)	**.83, .90**	.93, .99	.48, .58
CS	(1)	.877	**.960**	.740
	(2)	(.838, .916)	**(.928, .992)**	(.661, .819)
	(3)	.88	**.955**	.745
	(4)	.84, .90	**.94, .98**	.67, .81
AR	(1)	.466	.597	**.544**
	(2)	(.389, .542)	(.532, .662)	**(.485, .603)**
	(3)	.475	.595	**.55**
	(4)	.41, .53	.55, .67	**.47, .58**

Table 6.7

Simulation Results for n=20 and Med/Large Difference in Slope Parameters

Basis for: Analysis		Simulation: UN	Simulation: CS	Simulation: AR
			ρ=0.1	
UN	(1)	**.799**	.789	.745
	(2)	**(.733, .865)**	(.729, .849)	(.622, .868)
	(3)	**.80**	.79	.735
	(4)	**.75, .87**	.73, .83	.64, .89
CS	(1)	.763	**.773**	.747
	(2)	(.671, .854)	**(.700, .846)**	(.627, .867)
	(3)	.765	**.76**	.75
	(4)	.66, .78	**.73, .83**	.64, .88
AR	(1)	.717	.726	**.713**
	(2)	(.628, .806)	(.654, .798)	**(.589, .837)**
	(3)	.725	.72	**.705**
	(4)	.66, .78	.67, .80	**.60, .85**
			ρ=0.5	
UN	(1)	**.945**	.949	.756
	(2)	**(.903, .987)**	(.883, 1.02)	(.645, .867)
	(3)	**.94**	.95	.75
	(4)	**.92, .99**	.88, .99	.67, .82
CS	(1)	.953	**.958**	.827
	(2)	(.913, .992)	**(.904, 1.00)**	(.734, .921)
	(3)	.955	**.97**	.84
	(4)	.92, .99	**.90, .98**	.73, .89
AR	(1)	.798	.825	**.739**
	(2)	(.690, .906)	(.714, .936)	**(.637, .841)**
	(3)	.80	.825	**.72**
	(4)	.71, .89	.74, .89	**.67, .82**
			ρ=0.9	
UN	(1)	**1.00**	1.00	.999
	(2)	**(1.00, 1.00)**	(1.00, 1.00)	(.993, 1.00)
	(3)	**1.00**	1.00	1.00
	(4)	**1.00, 1.00**	1.00, 1.00	.99, 1.00
CS	(1)	1.00	**1.00**	.999
	(2)	(1.00, 1.00)	**(1.00, 1.00)**	(.993, 1.00)
	(3)	1.00	**1.00**	1.00
	(4)	1.00, 1.00	**1.00, 1.00**	.99, 1.00
AR	(1)	1.00	1.00	**.999**
	(2)	(1.00, 1.00)	(1.00, 1.00)	**(.993, 1.00)**
	(3)	1.00	1.00	**1.00**
	(4)	1.00, 1.00	1.00, 1.00	**.99, 1.00**

Table 6.8

Simulation Results for n=40 and No Difference in Slope Parameters

Basis for:		Simulation		
		UN	CS	AR
Analysis			ρ=0.1	
UN	(1)	**.083**	.068	.066
	(2)	**(.023, .143)**	(.018, .118)	(.013, .119)
	(3)	**.075**	.07	.06
	(4)	**.02, .12**	.03, .12	.03, .13
CS	(1)	.068	**.056**	.069
	(2)	(.008, .128)	**(.012, .010)**	(.024, .114)
	(3)	.07	**.055**	.065
	(4)	.01, .12	**.03, .10**	.03, .10
AR	(1)	.046	.037	**.059**
	(2)	(.002, .090)	(.007, .066)	**(.019, .099)**
	(3)	.045	.045	**.065**
	(4)	.01, .09	.02, .05	**.03, .09**
			ρ=0.5	
UN	(1)	**.070**	.072	.058
	(2)	**(.023, .117)**	(.032, .112)	(.014, .102)
	(3)	**.06**	.065	.06
	(4)	**.04, .12**	.05, .11	.03, .09
CS	(1)	.070	**.058**	.102
	(2)	(.042, .098)	**(.015, .101)**	(.061, .143)
	(3)	.07	**.06**	.10
	(4)	.05, .10	**.02, .10**	.07, .14
AR	(1)	.019	.015	**.044**
	(2)	(.002, .036)	(-.004, .034)	**(0.0, .094)**
	(3)	.02	.02	**.045**
	(4)	0.0, .03	0.0, .03	**.01, .08**
			ρ=0.9	
UN	(1)	**.058**	.058	.055
	(2)	**(.031, .085)**	(.010, .106)	(.017, .093)
	(3)	**.06**	.055	.06
	(4)	**.03 , .08**	.02, .10	.02, .08
CS	(1)	.067	**.051**	.155
	(2)	(.030, .104)	**(.001, .101)**	(.104, .206)
	(3)	.065	**.05**	.155
	(4)	.04, .11	**.01, .10**	.12, .20
AR	(1)	.005	0.0	**.044**
	(2)	(0.0, .019)	(0.0, 0.0)	**(.001, .087)**
	(3)	0.0	0.0	**.045**
	(4)	0.0, .02	0.0, 0.0	**.02, .08**

Table 6.9

Simulation Results for n=40 and Small Difference in Slope Parameters

Basis for Analysis		Simulation: UN	Simulation: CS	Simulation: AR
		ρ=0.1		
UN	(1)	**.454**	.412	.354
	(2)	**(.361, .547)**	(.334, .490)	(.261, .447)
	(3)	**.45**	.40	.345
	(4)	**.39, .55**	.35, .50	.28, .45
CS	(1)	.417	**.397**	.378
	(2)	(.317, .516)	**(.308, .486)**	(.281, .475)
	(3)	.405	**.39**	.385
	(4)	.36, .54	**.33, .49**	.28, .44
AR	(1)	.372	.349	**.350**
	(2)	(.284, .460)	(.274, .424)	**(.243, .457)**
	(3)	.365	.345	**.345**
	(4)	.33, .49	.29, .43	**.26, .43**
		ρ=0.5		
UN	(1)	**.643**	.652	.371
	(2)	**(.546, .740)**	(.515, .789)	(.255, .487)
	(3)	**.66**	.66	.37
	(4)	**.54, .70**	.55, .79	.29, .47
CS	(1)	.633	**.639**	.528
	(2)	(.546, .720)	**(.511, .767)**	(.448, .608)
	(3)	.645	**.63**	.515
	(4)	.55, .69	**.56, .79**	.47, .60
AR	(1)	.341	.342	**.379**
	(2)	(.261, .421)	(.249, .435)	**(.256, .502)**
	(3)	.33	.335	**.36**
	(4)	.28, .41	.28, .42	**.27, .47**
		ρ=0.9		
UN	(1)	**.989**	.999	.829
	(2)	**(.971, 1.00)**	(.993, 1.00)	(.767, .891)
	(3)	**.99**	1.00	.825
	(4)	**.98, 1.00**	.99, 1.00	.79, .89
CS	(1)	.992	**.999**	.935
	(2)	(.980, 1.00)	**(.993, 1.00)**	(.899, .971)
	(3)	.99	**1.00**	.935
	(4)	.98, 1.00	**.99, 1.00**	.91, .96
AR	(1)	.861	.948	**.821**
	(2)	(.787, .935)	(.918, .978)	**(.757, .885)**
	(3)	.87	.945	**.815**
	(4)	.80, .91	.93, .98	**.78, .88**

Table 6.10

Simulation Results for n=40 and Med/Large Difference in Slope Parameters

Basis for: Analysis		Simulation: UN	Simulation: CS	Simulation: AR
			ρ=0.1	
UN	(1)	**.975**	.971	.955
	(2)	**(.950, 1.00)**	(.923, 1.00)	(.909, 1.00)
	(3)	**.98**	.975	.96
	(4)	**.95, .99**	.92, 1.00	.92, .99
CS	(1)	.974	**.972**	.965
	(2)	(.949, .999)	**(.942, 1.00)**	(.915, 1.00)
	(3)	.975	**.98**	.975
	(4)	.95, .99	**.95, .99**	.92, .99
AR	(1)	.956	.965	**.962**
	(2)	(.916, .996)	(.927, 1.00)	**(.914, 1.00)**
	(3)	.955	.97	**.97**
	(4)	.92, .99	.93, .99	**.92, .99**
			ρ=0.5	
UN	(1)	**1.00**	1.00	.971
	(2)	**(1.00, 1.00)**	(1.00, 1.00)	(.948, .994)
	(3)	**1.00**	1.00	.97
	(4)	**1.00, 1.00**	1.00, 1.00	.95, .99
CS	(1)	1.00	**.999**	.986
	(2)	(1.00, 1.00)	**(.993, 1.00)**	(.964, 1.00)
	(3)	1.00	**1.00**	.98
	(4)	1.00, 1.00	**.99, 1.00**	.97, 1.00
AR	(1)	.994	.995	**.974**
	(2)	(.975, 1.00)	(.985, 1.00)	**(.940, 1.00)**
	(3)	1.00	.995	**.98**
	(4)	.97, 1.00	.99, 1.00	**.93, .99**
			ρ=0.9	
UN	(1)	**1.00**	1.00	1.00
	(2)	**(1.00, 1.00)**	(1.00, 1.00)	(1.00, 1.00)
	(3)	**1.00**	1.00	1.00
	(4)	**1.00, 1.00**	1.00, 1.00	1.00, 1.00
CS	(1)	1.00	**1.00**	1.00
	(2)	(1.00, 1.00)	**(1.00, 1.00)**	(1.00, 1.00)
	(3)	1.00	**1.00**	1.00
	(4)	1.00, 1.00	**1.00, 1.00**	1.00, 1.00
AR	(1)	1.00	1.00	**1.00**
	(2)	(1.00, 1.00)	(1.00, 1.00)	**(1.00, 1.00)**
	(3)	1.00	1.00	**1.00**
	(4)	1.00, 1.00	1.00, 1.00	**1.00, 1.00**

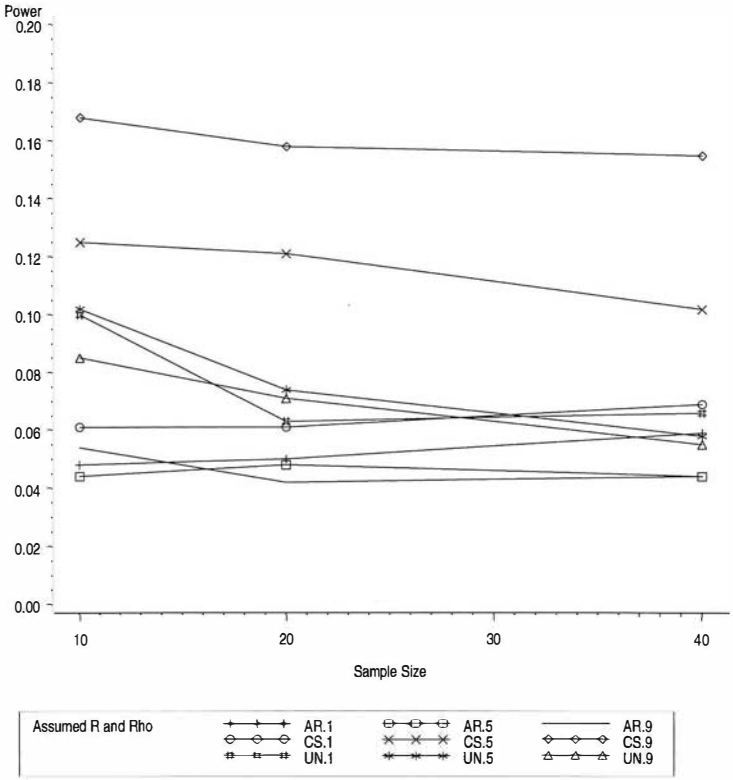

Figure 6.1: Plot of Simulated Power (Mean Value)

True **R**=AR and No Difference in Slope Parameters

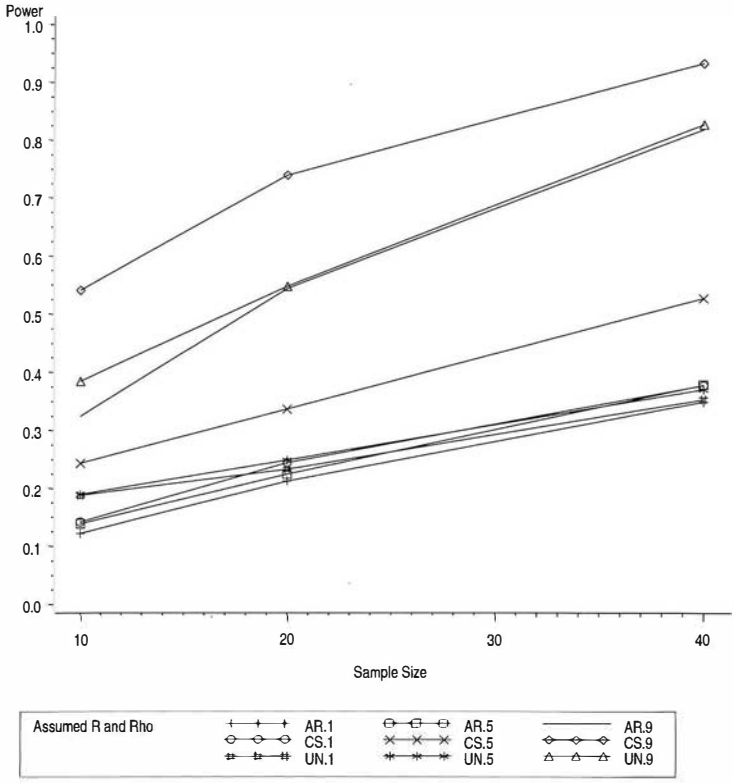

Figure 6.2: Plot of Simulated Power (Mean Value)

True **R**=AR and Small Difference in Slope Parameters

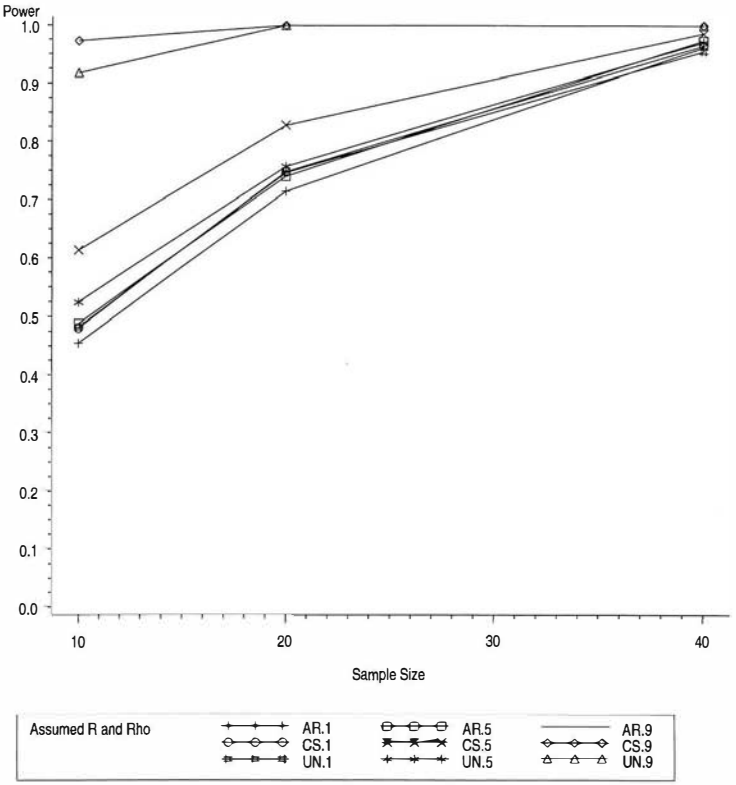

Figure 6.3: Plot of Simulated Power (Mean Value)

True **R**=AR and Med/Large Difference in Slope Parameters

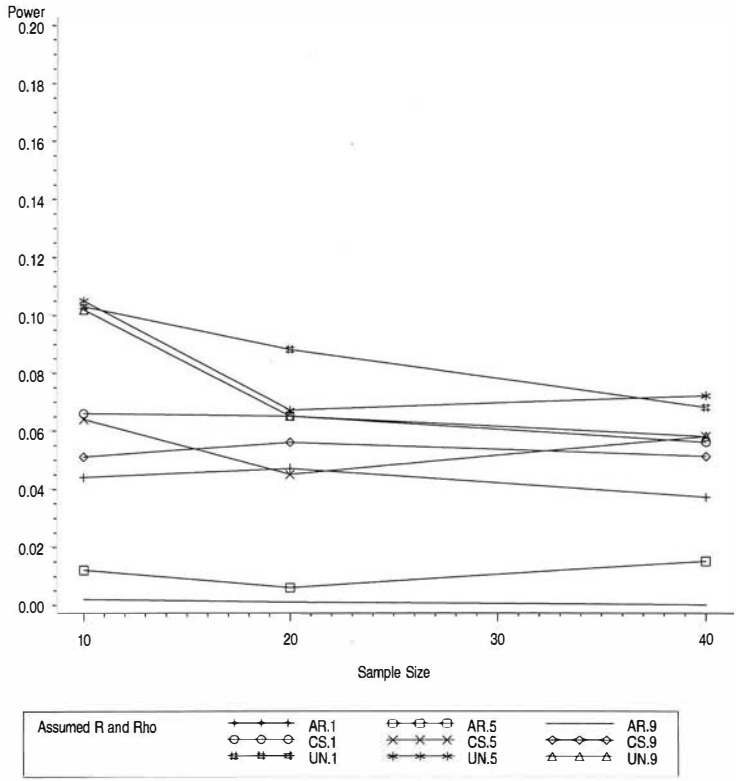

Figure 6.4: Plot of Simulated Power (Mean Value)

True **R**=CS and No Difference in Slope Parameters

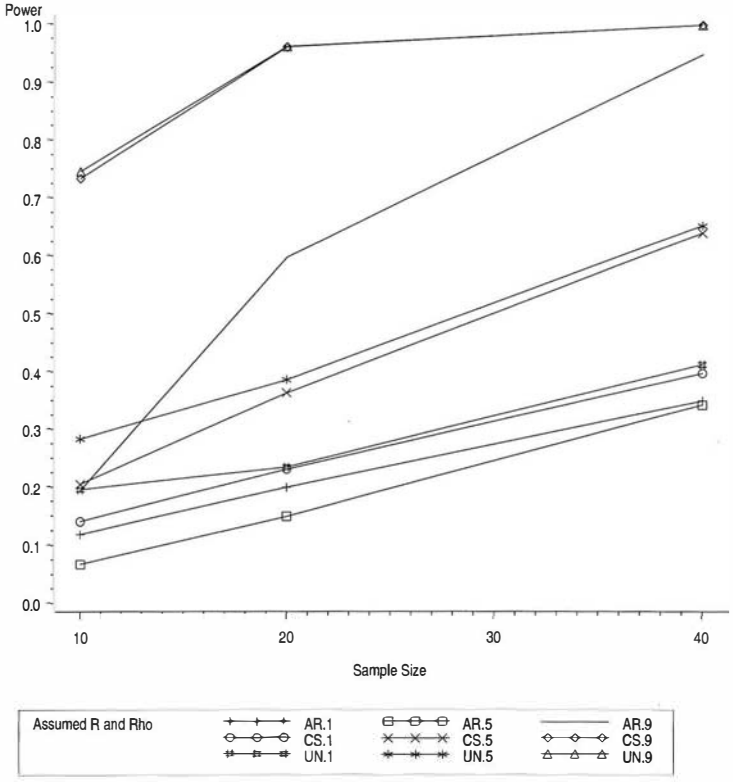

Figure 6.5: Plot of Simulated Power (Mean Value)

True **R**=CS and Small Difference in Slope Parameters

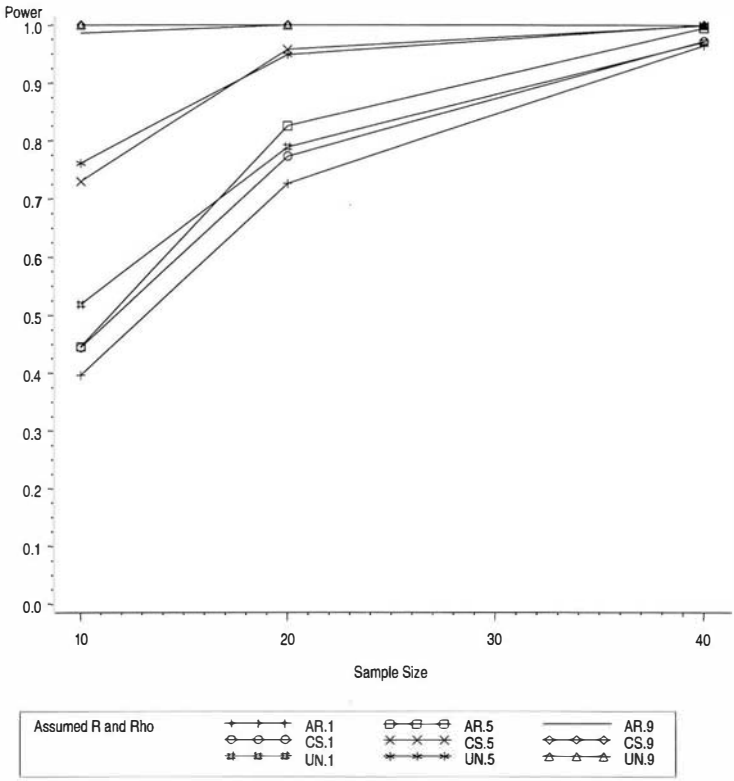

Figure 6.6: Plot of Simulated Power (Mean Value)

True **R**=CS and Med/Large Difference in Slope Parameters

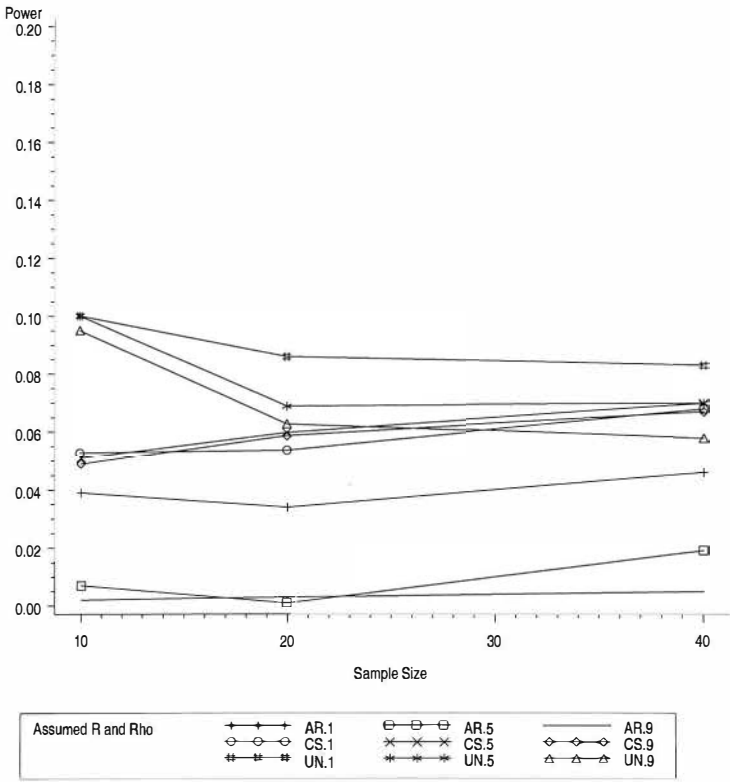

Figure 6.7: Plot of Simulated Power (Mean Value)

True **R**=UN and No Difference in Slope Parameters

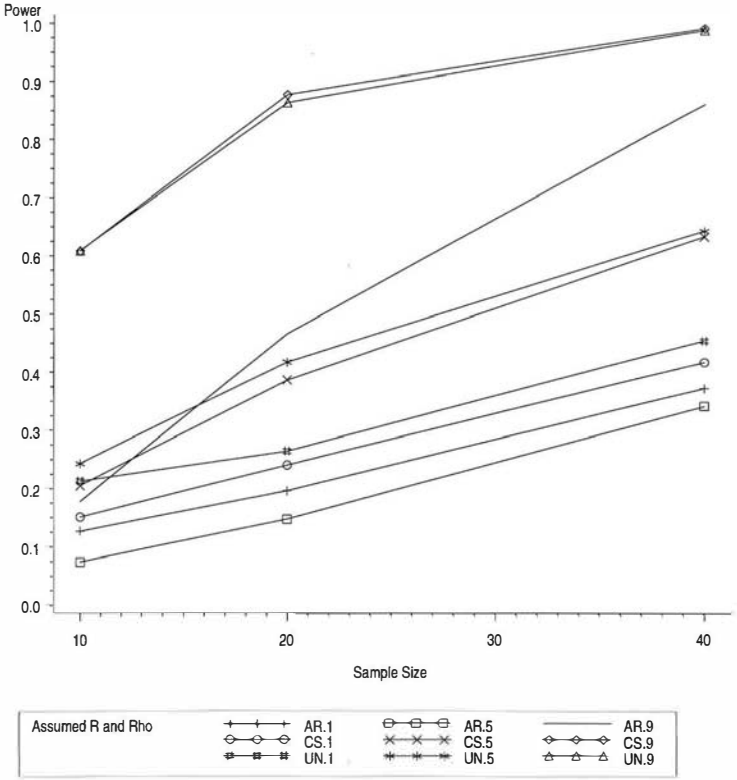

Figure 6.8: Plot of Simulated Power (Mean Value)

True **R**=UN and Small Difference in Slope Parameters

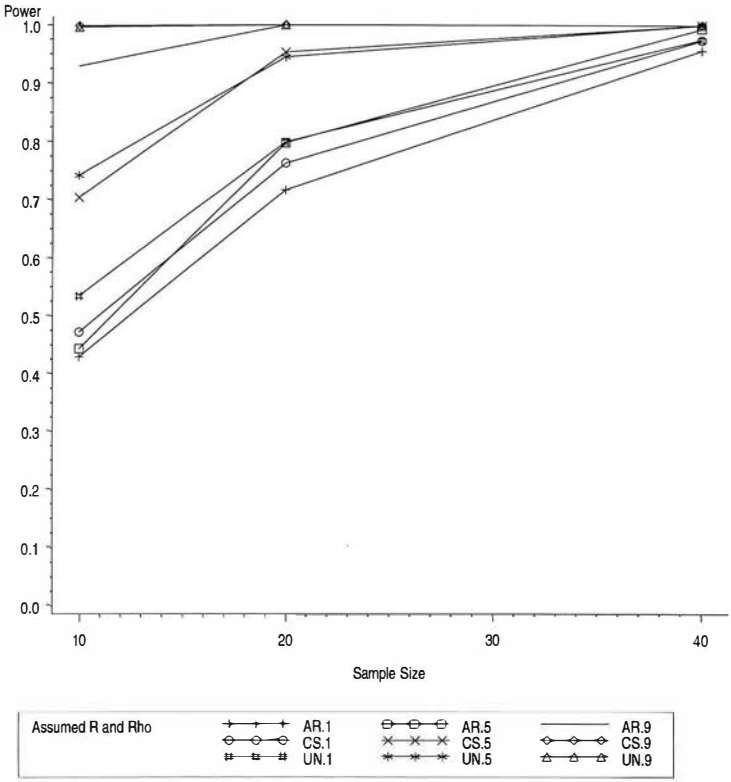

Figure 6.9: Plot of Simulated Power (Mean Value)

True **R**=UN and Med/Large Difference in Slope Parameters

6.4 Simulation Conclusions

The effect on power when misspecification of the covariance structure occurs is most noticeable when the true covariance structure is autoregressive i.e., power is inflated when assuming a covariance structure other than autoregressive. Also, misspecification by choosing an autoregressive covariance when the data does not follow an autoregressive covariance consistently underestimates the power. Power is greatly underestimated at higher levels of correlation. The only exception to the misspecification problem with autoregressive data is when modelling an unstructured covariance. Power estimates are close to those obtained assuming an autoregressive structure that has a high level of correlation; however, more parameters than necessary are being estimated with the unstructured covariance.

As the difference in slope parameters and sample sizes increase, misspecification of the covariance has a minimal effect on the power of the test of the fixed effects . At the largest levels of these combinations, powers greater that 0.95 are achieved for all tested covariance structures. It is conjectured that this relative stability in power is due to a ceiling effect, i.e., power must be less than or equal to 1.

Having discussed the effect of misspecification, attention is now turned toward the effects of underspecification. From the simulation study, adequate powers are obtained by assuming a compound symmetry covariance for unstructured data. In other words, there is essentially no difference between the power obtained for unstructured data when assuming a compound symmetry covariance as compared to the power that would be obtained with an unstructured covariance. For these cases, the compound symmetry power is slightly lower than the unstructured power but this difference decreases with an increase in sample size and correlation. However, the power of the test of fixed effects, when the data has an unstructured covariance, is greatly underestimated when assuming an autoregressive

covariance structure. Therefore, when assuming a covariance structure with fewer estimated parameters than the unstructured covariance, there appears to be only a small effect on the power when using a compound symmetry covariance, but when assuming an autoregressive covariance structure there is a large unwanted effect on power.

Based on this simulation, the following can be recommended when planning a power analysis for a longitudinal / repeated measures study in which testing of the fixed effects is the main interest. An autoregressive covariance structure should be used when the data is known to have an autoregressive structure. If it is known that the data does not follow an autoregressive covariance structure, then assuming a compound symmetry covariance may be sufficient. This fact is especially useful because many times estimates for all of the parameters of the unstructured covariance are not known. It is not recommended, however, to assume an autoregressive covariance structure when the data is actually unstructured.

From this simulation study, the objectives stated in Section 6.1 have been addressed. The first objective was to determine the effect of misspecification of the model's covariance structure on power. By seeing that an unwanted effect on power occurs when assuming a covariance structure other than autoregressive for autoregressive data, the first objective was addressed. The second objective was to investigate the effect of underspecification of the covariance structure. This objective was addressed by seeing that a compound symmetry, but not autoregressive structure, may be assumed for unstructured data.

Three concluding remarks are appropriate at this point. First, it must be remembered that the simulation focused on tests of the fixed effects only. None of these conclusions apply to situations when tests of the random effects or the variance parameters are also of interest.

Second, the covariance structures chosen for this simulation are considered as the three basic structured covariance models for longitudinal data (Grady and Helms, 1995). These structures may be adequate for some data analyses but may not be the best fitting for all data sets. Grady and Helms discuss some extensions of these basic covariance structures and how to determine which structure is best for the given data. They also state that if interest is mainly in the testing of the fixed effects, as it has been assumed here, then the basic covariance structures may be sufficient.

Finally, it should also be noted that these conclusions hold for longitudinal data which follow a Case II situation described in Chapter 4. It is reasonable to believe that if any other type of model, such as a Case I or III situation, was investigated, the conclusions may be different.

Chapter 7
Summary Comments and Future Works

7.1 Summary Comments

The primary focus of this research has been the development of a useful and efficient method to calculate power for the mixed linear model. The proposed method is an extension of the work done for the univariate and multivariate general linear models by O'Brien and Muller (1993). For the mixed linear model, power calculations are based on the approximate F test about the fixed effects proposed by Helms (1992). It is feasible to use this approximate F test to calculate power due to the test statistic's small sample properties (Helms and McCarroll, 1987 and 1991).

In Chapter 4, the noncentrality parameter is derived for three different cases. These cases are motivated by the different types of study designs used with the mixed linear model. Work by Helms (1992) is similar to the Case III situation described in Section 4.4; however, Helms makes the restrictive assumption that $\mathbf{R} = \sigma^2\mathbf{I}$. In this dissertation, the variance of the random error is allowed to take on any covariance structure and not just the simple covariance structure. In fact, no restriction is placed on the form of the covariance of the random effects or the random error for any of the three cases.

Even though the basic theory of the mixed linear model existed as early as Henderson's work in the early 1960's, the methods have not been widely applied. Reasons for the mixed model's relative obscurity are because interest has been mostly in the agriculture sciences and because of the lack of computing software. Within the last

decade, the mixed model has been used with increasing frequency. First, the mixed model has experienced wide spread attention across a broad spectrum of the statistical literature, in part due to the work by Laird and Ware (1982). Also, the availability of software such as Proc Mixed in SAS (SAS Institute, 1991) to implement mixed model methodology has led to an increased application. Included in this dissertation are the programs written in SAS/IML (SAS Institute, 1990) used for the calculating of mixed model power for each of the three cases discussed. These programs are easy to interpret and should be a useful tool for any statistician. They can also be used in conjunction with the data analysis in Proc Mixed. Power analysis, for a study in which the testing of the mixed model's fixed effects is the main research question, can be done using the programs listed in Appendix A.

7.2 Future Works

During the course of this research, several interesting ideas arose that warrant further investigation. In this dissertation, an approach to calculate power is developed for only the fixed effects portion of the mixed model. Random effects are typically treated as nuisance parameters; however, one is occasionally interested in testing hypotheses about the random effects. Examples for which tests of the random effects are common are studies which involve genetic and animal breeding applications (Henderson, 1984). Therefore, it may be useful to extend the method described in Chapter 4 to include a method for calculating power that includes random effects.

Secondly, there are some critics of the mixed model who suggest not using the mixed model when only the fixed effects are to be tested. As discussed in Chapter 3, one reason random effects are modelled is to allow for a reduction in the error variability. For these situations, it may be fruitful to determine if there is a level at which the variability being explained by the mixed model's random effects does not provide additional power

than one would have using the general linear model where the random effects are considered fixed.

As discussed in Section 6.4, the simulation only studied the three basic structured covariances used for longitudinal data. It may be of interest to repeat the simulation study and include extensions of these covariance structures that can be used to describe longitudinal data. The simulation can also be redone assuming a different study design, such as a Case I or Case III situation. The proposed simulation would determine if the conclusions drawn in Section 6.4 hold for all study designs or just for a balanced complete longitudinal study design.

Finally, throughout this dissertation, power for the mixed linear model was studied. Recently, nonlinear mixed effect models have been discussed in the literature (Pearson et al., 1994, Chinchilli, 1996, and Vonesh et al., 1996). Since the nonlinear mixed model is being used with increased frequency, it would be of further interest to extend the results found for the mixed linear model to include calculating power for the nonlinear mixed model.

List of References

List of References

Casella, G. and Berger, R.L. (1990). *Statistical Inference*. California: Wadsworth Inc.

Chinchilli, V.M. (1996). Personal Communication.

Drum, M.L. and McCullagh, P. (1993). REML estimation with exact covariance in the logistic mixed model. *Biometrics* **49**, 677-689.

Grady, J.J. and Helms, R.W. (1995). Model selection techniques for the covariance matrix for incomplete longitudinal data. *Statistics in Medicine* **14**, 1397-1416.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93-108.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320-340.

Harville, D.A. (1978). Alternative formulations and procedures for the two-way mixed model. *Biometrics* **34**, 441-453.

Helms, R.W. (1992). Intentionally incomplete longitudinal designs: methodology and comparison of some full span designs. *Statistics in Medicine* **11**, 1889-1913.

Helms, R.W. and McCarroll, K. (1991). Comparisons of approximate F statistics and their small sample distributions for fixed effects hypotheses in mixed models. Manuscript submitted for publication.

Henderson, C.R. (1984). *Applications of Linear Models in Animal Breeding*. Canada: University of Guelph.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447.

Henderson, C.R., Kempthorne, O., Searle, S.R., and von Krosigk, C.M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192-218.

Hicks, C.R. (1973). *Fundamental Concepts in the Design of Experiments*. New York: Holt, Rinehart, and Winston.

Hsuan, F.C. (1993). Estimating treatment means in mixed effect ANOVA model for bioequivalence studies. *Biometrics* **49**, 703-713.

Jennrich, R.I. and Schlucter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805-820.

Jeske, D.R. and Harville, D.A. (1988). Prediction-interval procedures and (fixed-effects)confidence-interval procedures for mixed linear models. *Communication in Statistics - Theory and Methods* **17**(4), 1053-1087.

Kacker, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79**, 853-962.

Kimball, A.W. (1957). Errors of the third kind in statistical consulting. *Journal of the American Statistical Association* **52**, 133-142.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Lee, Y. (1971). Asymptotic formulae for the distribution of a multivariate test statistic: power comparisons of certain multivariate tests. *Biometrika* **58**, 647-651.

Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS® System for Mixed Models*. Cary, NC: SAS Institute Inc.

McCarroll, K. and Helms, R.W. (1987). An evaluation of some approximate F statistics and their small sample distributions for the mixed model with linear covariance structure. The Institute of Statistics Mimeo Series, Department of Biostatistics, University of North Carolina, Chapel Hill.

McLean, R.A. and Sanders. W.L. (1988). Approximating degrees of freedom for standard errors in mixed linear models. *American Statistical Association 1988 Proceedings of the Statistical Computing Section.* 50-64.

McLean, R.A., Sanders. W.L., and Stroup, W.W. (1991). A unified approach to mixed linear models. *The American Statistician* **45**, 54-64.

Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* **41**, 153-165.

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

Muller, K.E., La Vange, L.M., Ramey, S.L., and 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'Brien, R.G. and Muller, K.E. (1993). Unified power analysis for t-tests through multivariate hypotheses in *Applied Analysis of Variance in Behavioral Science* ed. Lynne K. Edwards. New York: Marcel Dekker Inc. 297-344.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-554.

Pearson, J.D., Morrell, C.H., Landis, P.K., Carter, H.B., and Brant, L.J. (1994). Mixed-effects regression models for studying the natural history of prostate disease. *Statistics in Medicine* **13**, 587-601.

Rao. C.R. (1973). *Linear Statistical Inference and Its Applications* 2nd ed. New York: John Wiley & Sons.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science* **6**, 15-51.

SAS Institute Inc. (1990). *SAS/IML Software®: Usage and Reference*, Version 6, First Edition. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1992). *SAS® Technical Report P-229: SAS/STAT® Software: Changes and Enhancements*. Release 6.07. Cary, NC: SAS Institute Inc.

Schaeffer, L.R. and Wilton, J.W. (1978). Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* **34**, 199-208.

Searle, S.R. (1971). *Linear Models*. New York: John Wiley & Sons.

Sherrull, D.L., Enright, P.L., Lebowitz, M.D., and Burrows, B. (1994). Longitudinal analysis of the effects of smoking onset and cessation on pulmonary function. *American Journal of Respiratory Critical Care Medicine* **149**, 591-597.

Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *The Annals of Mathematical Statistics* **40**(3), 942-952.

Vonesh, E.F., Chinchilli, V.M., and Pu, K. (1996). Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics* **52**, 572-587.

Wolfinger, R. (1992). A tutorial on mixed models. SAS Institute Inc., Cary, North Carolina.

Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in Statistics-Simulation* **22**(4), 1079-1106.

Appendix A

Programs for Examples

```
*************************************************************************
This program in Proc IML of SAS conducts a mixed model power analysis for a Case I
design.

The user must construct the following matrices and scalar values:
        x_e the (Ne × p) fixed effects essence matrix
        z_e the (Ne × q) random effects essence matrix
        w the (Ne × Ne) weight matrix
        g the (q × q) random effects variance covariance matrix
        r_e the (Ne × Ne) random error variance covariance essence matrix
        beta the (p × 1) fixed effects parameter estimates
        kprime the (k × p) contrast matrix of the fixed effects that is of interest
        rankxz the rank of the [X Z] matrix
        alpha the probability of type I error
        _i1_, _12_, and _by_ the starting, stopping and by values for N, the total sample
                size used for calculating power.  Note: (_i1_ > rankxz).

A dataset containing N and its corresponding power can be created from the matrix matp
which is made in the IML function.
*************************************************************************;

proc iml;

start power1;
        kpbeta=kprime*beta;
        df_n=nrow(kprime);
        _xpw_=x_e`*w;
        _zgzpw_=z_e*g*z_e`*w;
        do _n_= _i1_ to _i2_ by _by_;
                inv1=inv(_n_#_zgzpw_ + r_e);
                inv2=inv(_xpw_*inv1*x_e);
                inv3=inv(kprime*inv2*kprime`);
                lambda_s=kpbeta`*inv3*kpbeta;
                lambda=_n_#lambda_s;
                df_d=_n_ - rankxz;
                fcrit=finv(1-alpha, df_n, df_d, 0);
                power=1-probf(fcrit,df_n, df_d, lambda);
```

```
                outp=_n_ || power;
                matp=matp // outp;
        end;
finish power1;

*** Matrices defined for Case I Example 1 ***;
        x_e={1 0,
             0 1,
             1 0,
             0 1,
             1 0,
             0 1};

        zstar={1 1 0,
               1 0 1};
        z_e=I(3)@zstar;

        gstar={.1073 .0510 .0510};
        gstar=Diag(gstar);
        g=I(3)@gstar;

        rstar=.0485#I(2);
        r_e=I(3)@rstar;

        beta={50.95,
              51.96};

        kprime={1 -1};

        w=(1/6)#I(6);

        rankxz=6;

        _i1_=8;   _i2_=42;   _by_=2;

        alpha=.05;
        run power1;
        coln={'N', 'Power'};
        create powera from matp[colname=coln];
        append from matp;
        free outp matp;

        alpha=.01;
        run power1;
        coln={'N', 'Power'};
        create powerb from matp[colname=coln];
        append from matp;

proc print noobs data=powera;
        title2'Case1 Ex1';
        title3'Power for alpha=.05';
```

```
proc print noobs data=powerb;
        title2'Case1 Ex1';
        title3'Power for alpha=.01';

endsas;


*** Matrices defined for Case I Example 2 ***;
        xstar=I(3);
        x_e=xstar//xstar//xstar//xstar;

        z_e=I(12);

        r_e=3.3#I(12);

        w=(1/12)#I(12);

        beta={ 10.31,
               16.09,
               19.37};

        kprime={1 -1  0,
                1  0 -1};

        alpha=.05;

        rankxz=12;

        _i1_=24;   _i2_=480;   _by_=12;

        gstar={.42 26.7 56.94};
        gstar=Diag(gstar);
        g=I(4)@gstar;

        run power1;
        coln={'N',  'Power'};
        create powera from matp[colname=coln];
        append from matp;
        free outp matp;

        g=26.8#I(12);

        run power1;
        coln={'N',  'Power'};
        create powerb from matp[colname=coln];
        append from matp;


proc print noobs data=powera;
        title2'Case1 Ex2';
```

```
        title3'Power for Un(1) G';

proc print noobs data=powerb;
        title2'Case1 Ex2';
        title3'Power for Simple G';
```

```
*************************************************************************
```
This program in Proc IML of SAS conducts a mixed model power analysis for a Case II or III design.

The user must construct the following matrices and scalar values:

- x_e the (Ne × p) fixed effects essence matrix
- z_e the (Ne × cT) random effects essence matrix
- w the (Ne × Ne) weight matrix
- g_e the (cT × cT) random effects variance covariance essence matrix
- r_e the (Ne × Ne) random error variance covariance essence matrix
- beta the (p × 1) fixed effects parameter estimates
- kprime the (k × p) contrast matrix of the fixed effects that is of interest
- rankxz the rank of the [X Z] matrix
- alpha the probability of type I error
- _i1_, _12_, and _by_ the starting, stopping and by values for N, the total number of observations used for calculating power. Note: (_i1_ > rankxz).

A dataset containing N, the total number of observations, and its corresponding power can be created from the matrix matp which is made in the IML function.

```
*************************************************************************;

proc iml ;

start power2_3;
        kpbeta=kprime*beta;
        df_n=nrow(kprime);
        inv1=inv(z_e*g_e*z_e` + r_e);
        inv2=inv(x_e`*w*inv1*x_e);
        inv3=inv(kprime*inv2*kprime`);
        lambda_s=kpbeta`*inv3*kpbeta;
        do _n_= _i1_ to _i2_ by _by_;
                lambda=_n_#lambda_s;
                df_d=_n_ - rankxz;
                fcrit=finv(1-alpha, df_n, df_d, 0);
                power=1-probf(fcrit,df_n, df_d, lambda);
                outp=_n_ || power;
                matp=matp // outp;
        end;
finish power2_3;

**** Matrices defined for Case2 Example1****;
        x_e={1 -1 0  0,
             1  0 0  0,
             1  1 0  0,
             0  0 1 -1,
             0  0 1  0,
             0  0 1  1};

        z_e={1 0,
             1 0,
```

```
            1 0,
            0  1,
            0  1,
            0  1};

        w=(1/6)#I(6);

        gstar={3.0};
        g_e=I(2)@gstar;

        rstar={2.4   -.2   .9,
                -.2   1.2   .01,
                .9    .01   3.5};
        r_e=I(2)@rstar;

        kprime={1 0 -1  0,
                0 1  0 -1};

        beta={21.2,
              1.4,
              20.3,
              .95};

        rankxz=4;

        _il_=12;   _i2_=900;   _by_=12;

        alpha=.05;
        run power2_3;
        coln={'N',  'Power'};
        create powera from matp[colname=coln];
        append from matp;
        free outp matp;

        alpha=.01;
        run power2_3;
        coln={'N',  'Power'};
        create powerb from matp[colname=coln];
        append from matp;

proc print noobs data=powera;
        title1'Case2 Ex1';
        title2'Power for alpha=.05';
        title3'N=total number of observations';
        title4'number of subjects is N/3';

proc print noobs data=powerb;
        title1'Case2 Ex1';
        title2'Power for alpha=.01';
        title3'N=total number of observations';
        title4'number of subjects is N/3';
```

```
endsas;


****Matrices defined for Case2 Example2 ****;
        x_e={1 0,
             1 0,
             1 0,
             0 1,
             0 1,
             0 1};

        z_e=x_e;

        w=(1/6)#I(6);

        gstar={325.33 328.58,
               328.58 321.32};
        g_e=gstar;

        rstar={42.36 42.36 42.36 42.31 42.31 42.31};
        rstar=diag(rstar);
        r_e=rstar;

        kprime={1 -1};

        beta={123.5,
              120.47};

        rankxz=2;

        _i1_=6;   _i2_=510;   _by_=12;

        alpha=.05;
        run power2_3;
        coln={'N', 'Power'};
        create powera from matp[colname=coln];
        append from matp;
        free outp matp;

        alpha=.01;
        run power2_3;
        coln={'N', 'Power'};
        create powerb from matp[colname=coln];
        append from matp;

proc print noobs data=powera;
        title1'Case2 Ex2';
        title2'Power for alpha=.05';
        title3'N=total number of observations';
        title4'number of subjects is N/6';
```

```
proc print noobs data=powerb;
        title1'Case2 Ex2';
        title2'Power for alpha=.01';
        title3'N=total number of observations';
        title4'number of subjects is N/6';

endsas;


****Matrices defined for Case3 Example1****;
        x_e={1 1 0 0,
             1 2 0 0,
             1 3 0 0,
             1 4 0 0,
             0 0 1 1,
             0 0 1 2,
             0 0 1 3,
             0 0 1 4,
             1 1 0 0,
             1 2 0 0,
             1 3 0 0,
             0 0 1 1,
             0 0 1 2,
             0 0 1 3,
             1 1 0 0,
             1 2 0 0,
             0 0 1 1,
             0 0 1 2};

        z4={1,1,1,1};
        z3={1,1,1};
        z2={1,1,};
        z_e=Block(z4, z4, z3, z3, z2, z2);

        gstar={.39};
        g_e=I(6)@gstar;

        rstarm={1905  20.8  .23  .002,
                20.8  1905  20.8  .23,
                .23   20.8  1905  20.8,
                .002  .23   20.8  1905};
        rstar3=rstarm[{1 2 3}. {1 2 3}];
        rstar2=rstarm[{1 2}. {1 2}];
        r_e=Block(rstarm, rstarm, rstar3, rstar3, rstar2, rstar2);

        kprime={1 0 -1 0,
                0 1 0 -1};

        beta={-5,
              55.5,
```

```
            -4.5,
            52.3};

        rankxz=8;

        alpha=.05;

        _i1_=18;   _i2_=3690;   _by_=36;

        w4=(1/18)#I(4);
        w3=(1/18)#I(3);
        w2=(1/18)#I(2);
        w=Block(w4, w4, w3, w3, w2, w2);

        run power2_3;
        coln={'N', 'Power'};
        create powera from matp[colname=coln];
        append from matp;
        free outp matp;

        w4=(3/40)#I(4);
        w3=(2/40)#I(3);
        w2=(1/40)#I(2);
        w=Block(w4, w4, w3, w3, w2, w2);

        run power2_3;
        coln={'N', 'Power'};
        create powerb from matp[colname=coln];
        append from matp;

proc print noobs data=powera;
        title1'Case3 Ex1';
        title2'Power for Equal SS';
        title3'N=total number of observations';

proc print noobs data=powerb;
        title1'Case3 Ex1';
        title2'Power for 3:2:1';
        title3'N=total number of observations';

endsas;

****Matrices defined for Case3 Example2****;
        x1={1 -2 0 0 0 0,
            1 -1 0 0 0 0,
            1  0 0 0 0 0,
            1  1 0 0 0 0,
            1  2 0 0 0 0};
        x2={0 0 1 -2 0 0,
            0 0 1 -1 0 0,
            0 0 1  0 0 0,
```

```
      0 0 1  1 0 0,
      0 0 1  2 0 0};
x3={0 0 0 0 1 -2 ,
      0 0 0 0 1 -1,
      0 0 0 0 1  0,
      0 0 0 0 1  1,
      0 0 0 0 1  2};
x4={1 -2 0 0 0 0,
      1  0 0 0 0 0,
      1  2 0 0 0 0};
x5={0 0 1 -2 0 0,
      0 0 1  0 0 0,
      0 0 1  2 0 0};
x6={0 0 0 0 1 -2 ,
      0 0 0 0 1  0,
      0 0 0 0 1  2};
x_e=x1//x2//x3//x4//x5//x6;

z1={1 -2,
      1 -1,
      1  0,
      1  1,
      1  2};
z2=z1;
z3=z1;
z4={1 -2,
      1  0,
      1  2};
z5=z4;
z6=z4;
z_e=Block(z1, z2, z3, z4, z5, z6);

g_star={1.15   .163,
          .163   .039};
g_e=I(6)@g_star;

r_e=.1255#I(24);

kprime={0 1 0 -1 0 0,
          0 1 0 0 0 -1};

beta={-1.39,
        -.035,
        -2.39,
        -.176,
        -3.38,
        -.318};

rankxz=12;

alpha=.05;
```

```
        _i1_=24;   _i2_=600;   _by_=6;

        w=(1/24)#I(24);

        run power2_3;
        coln={'N',  'Power'};
        create powera from matp[colname=coln];
        append from matp;
        free outp matp;

        w1=(1/42)#I(15);
        w2=(3/42)#I(9);
        w=Block(w1, w2);

        run power2_3;
        coln={'N',  'Power'};
        create powerb from matp[colname=coln];
        append from matp;

proc print noobs data=powera;
        title1'Case3 Ex2';
        title2'Power for Equal SS';
        title3'N=total number of observations';

proc print noobs data=powerb;
        title1'Case3 Ex2';
        title2'Power for 1:4';
        title3'N=total number of observations';


endsas;
```

Appendix B

Simulation Study Program

```
*************************************************************************
This program simulates Normal (XB,R) data, analyzes the data in Proc Mixed, performs
the contrast of interest, determines whether the hypothesis is rejected and then calculates the
empirical power for each of the three variance-covariance structures being tested.

For each case being simulated, the following must be supplied:
        in the Proc IML statement,    n- the number of subjects per group,
                                      sigma_pt- the R matrix for an individual, and
                                      beta- in the form int1 slope1 int2 slope2
        in the data two statement,    total- the number of subjects per group and
                                      Title statements desribing the case being simulated
*************************************************************************;


options ls=80 ps=56 nonotes;
title3'Simulation for #### total subjects per group';
title4' with #### difference in the slope parameters';
title5'Assuming #### var-cov structure with correlation of####';


%macro simu;
proc iml;

        n=10; *n=20; *n=40;                 /*number subjects per group*/
        z=normal(repeat(0,n,8));
        *print z;

        sigma_pt={1.5 .15 .15 .15,
                  .15 1.5 .15 .15,
                  .15 .15 1.5 .15,
                  .15 .15 .15 1.5};         /*Var-cov for CS and correlation=.1*/

        sigma_pt={1.5 .75 .75 .75,
                  .75 1.5 .75 .75,
                  .75 .75 1.5 .75,
                  .75 .75 .75 1.5};         /*Var-cov for CS and correlation=.5*/

        sigma_pt={1.5 1.35 1.35 1.35,
```

```
            1.35 1.5 1.35 1.35,
            1.35 1.35 1.5 1.35,
            1.35 1.35 1.35 1.5};        /*Var-cov for CS and correlation=.9*/

sigma_pt={1.5 .15 .015 .0015,
           .15 1.5 .15 .015,
          .015 .15 1.5 .15,
        .0015 .015 .15 1.5};            /*Var-cov for AR and correlation=.1*/

sigma_pt={1.5 .75 .375 .1875,
           .75 1.5 .75 .375,
          .375 .75 1.5 .75,
        .1875 .375 .75 1.5};            /*Var-cov for AR and correlation=.5*/

sigma_pt={1.5 1.35 1.215 1.0935,
          1.35 1.5 1.35 1.215,
         1.215 1.35 1.5 1.35,
       1.0935 1.215 1.35 1.5};          /*Var-cov for AR and correlation=.9*/

sigma_pt={    1 .115 .129 .141,
           .115 1.33 .149 .163,
           .129 .149 1.66 .182,
           .141 .163 .182   2};         /*Var-cov for UN and correlation=.1*/

sigma_pt={    1 .577 .644 .707,
           .577 1.33 .743 .815,
           .644 .743 1.66 .911,
           .707 .815 .911   2 };        /*Var-cov for UN and correlation=.5*/

sigma_pt={    1 .967 1.16 1.27,
           .967 1.33 1.34 1.47,
          1.16 1.34 1.66 1.64,
          1.27 1.47 1.64   2};          /*Var-cov for UN and correlation=.9*/


sigma=Block(sigma_pt, sigma_pt);
shalf=root(sigma);

beta={4.2, 1.25, 4.95, 1.25};           /*for equal slopes*/
beta={4.2, 1.25, 4.95, 1.45};           /*for small difference in slopes*/
beta={4.2, 1.25, 4.95, 1.70};           /*for med/large difference in slope*/;

x={1 1 0 0,
   1 2 0 0,
   1 3 0 0,
   1 4 0 0,
   0 0 1 1,
   0 0 1 2,
   0 0 1 3,
   0 0 1 4};
xbeta=X*beta;
```

```
*print xbeta;
j=j(n,1,1);
y=z*shalf + j*xbeta`;
*print y;

person=1:n;

c={'person', 'y11', 'y12', 'y13', 'y14', 'y21', 'y22', 'y23','y24'};

sim=person`lly;
create simu from sim[colname=c];
append from sim;

data two;
        set simu;
        total=10; total=20; total=40;   /*number of subjects per group*/
        y=y11; time=1; group=1; output;
        y=y12; time=2; group=1; output;
        y=y13; time=3; group=1; output;
        y=y14; time=4; group=1; output;
        y=y21; time=1; group=2; person=person+total; output;
        y=y22; time=2; group=2; person=person; output;
        y=y23; time=3; group=2; person=person; output;
        y=y24; time=4; group=2; person=person; output;
        drop total y11 y12 y13 y14 y21 y22 y23 y24;

*proc print data=two;

%global _disk_;
%let _disk_=on;
%global _print_;
%let _print_=off;

proc mixed data=two;
        class person group;
        model y=group time*group/noint;
        repeated/ type=un subject=person;
        contrast 'Slope test un' time*group 1 -1;
        make 'Contrast' out=test_un(rename=(P_F=p_un));

proc mixed data=two;
        class person group;
        model y=group time*group/noint;
        repeated/ type=cs subject=person;
        contrast 'Slope test cs' time*group 1 -1;
        make 'Contrast' out=test_cs(rename=(P_F=p_cs));

proc mixed data=two;
        class person group;
        model y=group time*group/noint;
        repeated/ type=ar(1) subject=person;
```

```
            contrast 'Slope test ar' time*group 1 -1;
            make 'Contrast' out=test_ar(rename =(P_F=p_ar));

        data tests (drop=source ndf ddf f);
            merge test_un test_cs test_ar;

        proc datasets nolist;
            delete simu two test_un test_cs test_ar;
%mend;


%macro set(s);
        data total;
            %do _i_=1 %to &s;
                %simu;
                data total;
                set total tests;
                proc datasets nolist;
                delete tests;
            %end;

        data total;
        set total;
        if _n_=1 then delete;
        rject_un=0; rject_cs=0; rject_ar=0;
        if p_un < .05 then rject_un=1;
        if p_cs < .05 then rject_cs=1;
        f p_ar < .05 then rject_ar=1;

        *proc print data=total;

        proc means mean data=total noprint;
            var rject_un rject_cs rject_ar;
            output out=pow mean=power_un power_cs power_ar;

        proc datasets nolist;
            delete total;
        *proc print data=pow;
%mend;


%macro runs(r);
        data power;

        %do _r_=1 %to &r;
            %set(100);
            data power;
            set power pow;
        %end;

        data power;
```

```
            set power;
            if _n_=1 then delete;
        proc print data=power;

        proc means data=power;
            var power_un power_cs power_ar;

        proc univariate plot data=power;
            var power_un power_cs power_ar;

        proc datasets nolist kill;
%mend;

%runs(10);
```

Vita