FIXED EFFECTS INFERENCE FOR CLUSTERED DATA IN GAUSSIAN LINEAR MODELS

Jacqueline L. Johnson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics, School of Public Health.

Chapel Hill
2007

Approved by:

Co-Advisor: Diane J. Catellier

Co-Advisor: Keith E. Muller

Reader: Lisa M. LaVange

Reader: David M. Murray

Reader: John S. Preisser

ABSTRACT

Jacqueline L. Johnson: Fixed Effects Inference for Clustered Data In Gaussian Linear Models
(Under the direction of Dr. Diane J. Catellier and Dr. Keith E. Muller)

Important public health research often requires the use of community based studies due to logistical, ethical and cost constraints. Such designs require special methods of analysis. Gaussian clustered data are often analyzed with either a mixed effects linear model on individual level data or two-stage analysis of cluster means. For data with a large number of clusters and large number of observations within each cluster, both techniques provide unbiased hypothesis tests. In small samples with unbalanced data, however, even moderate imbalance in cluster size across treatment groups can bias hypothesis tests in the two stage analysis of cluster means. The use of large sample approximations for one-stage mixed model test statistics for analysis of small, unbalanced clustered experiments may also lead to inaccurate hypothesis tests.

I derived a formulation of quadratic form theory which leads to a method to obtain exact test size for hypothesis tests in the two stage model. This theory is used in an enumeration study of type I error for a test of treatment difference in the two stage analysis of cluster means where means are either unweighted or weighted by their cluster size. These enumerations focus on scenarios of imbalance common to non-randomized cluster data settings.

Next I performed a simulation study of type I error for a test of treatment difference in both the analysis of individual level data and of cluster means for scenarios of imbalance common to randomized clustered data trials. Ten methods were considered; of these, a two stage analysis of cluster means with means weighted by their theoretical variance controlled type I error under the most cases. In this analysis, the weights contain restricted maximum likelihood estimates of variance components estimated from the individual level data and are constrained to be positive.

Many current clustered data studies currently show a misalignment between power calculations and data analysis; that is, the power analysis is done for a simplified version of the

actual test computed. I showed how to perform an appropriate and valid power analysis for the previous two stage method and applied this to a study on adolescent drinking behavior.

ACKNOWLEDGMENTS

First, I would like to thank my advisors, Diane Catellier and Keith Muller, for their mentorship and friendship throughout the writing of this dissertation. I would also like to thank Keith Muller for his leadership, caring, and encouragement throughout all stages of my graduate education. I am also grateful to the members of my committee, Lisa LaVange, David Murray, and John Preisser, for their helpful ideas and comments. I would also like to thank Robert Hamer, Michael Schell, Larry Kupper, and Ruth Marinshaw for their guidance and friendship during specific phases of my education. Finally, I am very grateful to my family and friends for their continual love and support, without which I would have never attempted nor finished this dissertation and doctoral program.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES ix

1 Introduction and Literature Review **1**

1.1 Introduction 1

1.2 Aims 3

1.3 Notation 5

1.4 Statement of Models and Hypothesis 8

1.5 Literature Review 10

1.6 Motivating Data Example 16

2 Exact Type I Error in the Two Stage Analysis of Cluster Means **20**

2.1 Introduction 20

2.2 Hypothesis Testing for Cluster Means 22

2.3 Theoretical Result to Compute Probabilities Under Violation of Assumptions . . 26

2.4 Description of Enumerations 27

2.5 Results of Enumeration Study 29

2.6 Conclusions and Recommendations 32

2.7 Proof 33

3 Comparison of Type I Error for One Stage and Two Stage Models **43**

3.1 Introduction 43

3.2 Statement of Models and Hypothesis 44

3.3 Hypothesis Testing for Clustered Data with Balanced Cluster Sizes 47

3.4 Hypothesis Testing for Clustered Data with Unbalanced Cluster Sizes 51

3.5 Description of Simulations 53

3.6 Results of Simulation Study 57

3.7 Conclusions 60

4 Power Analysis for Continuing a Longitudinal Cluster Sample **80**

4.1 Introduction 80

4.2 Literature Review 81

4.3 Performing the Power Analysis 84

4.4 Data Example 88

4.5 Further Remarks 89

5 Summary and Future Research **91**

Bibliography 94

LIST OF TABLES

1.1 Notation in This Document Compared to Traditional Mixed Model Notation . . 7

1.2 Summary of Non Matrix Notation 7

1.3 Type I Error for One Stage and Two Stage Analyses 19

2.1 Summary of Non Matrix Notation 34

2.2 Type I error over all cases and by ρ, $m_1 \times m_2$ and m_2/m_1 35

2.3 Type I error by $\bar{n}_1 \times \bar{n}_2$ and $\bar{n}_2/\bar{n}_1$ 36

2.4 Type I error by $r_1 \times r_2$ and r_2/r_1 37

2.5 Type I error by $m_2/m_1 \times \bar{n}_2/\bar{n}_1$ 38

2.6 Type I error by $m_2/m_1 \times \bar{n}_2/\bar{n}_1$ (cont.) 39

3.1 Number of convergent scenarios for tests 2, 4, and 10, by ρ and $m_1 \times m_2$ 62

3.2 Number of scenarios with a negative variance component estimate, by ρ and $m_1 \times m_2$ 63

3.3 Type I error over all cases 64

3.4 Number (%) of total (N=9,090) cases where test i (down) is biased and test j (across) is unbiased. 65

3.5 Difference in Number (i, j) and Number (j, i) from Table 3.4. 65

3.6 Type I error by ρ 66

3.7 Type I error by $m_1 + m_2$ 67

3.8 Type I error for selected scenarios of balance 68

3.9 Type I error for selected scenarios of balance (cont.) 69

LIST OF FIGURES

2.1 Type I Error for $\rho \times m_1 \times m_2 \times \bar{n}_1 \times \bar{n}_2$ - Analysis of Unweighted Means 40

2.2 Type I Error for $\rho \times m_1 \times m_2 \times \bar{n}_1 \times \bar{n}_2$ - Analysis of Means Weighted By Cluster Size 41

2.3 Type I Error for $\rho \times m_1 \times m_2 \times \bar{n}_1 \times \bar{n}_2$ - Comparison of Weights 42

3.1 Type I Error for Test 1 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 70

3.2 Type I Error for Test 2 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 71

3.3 Type I Error for Test 3 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 72

3.4 Type I Error for Test 4 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 73

3.5 Type I Error for Test 5 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 74

3.6 Type I Error for Test 6 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 75

3.7 Type I Error for Test 7 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 76

3.8 Type I Error for Test 8 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 77

3.9 Type I Error for Test 9 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 78

3.10 Type I Error for Test 10 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$ 79

4.1 Power as a Function Mean Difference 90

Chapter 1

Introduction and Literature Review

1.1 Introduction

The term "clustered data" commonly refers to data collected on individuals who are nested within a specific geographical or civil unit, e.g., children within schools, employees within worksites, or patients within physician practices. Clustered designs are often intentionally used to study the relationship of characteristics at the individual and cluster level on the response of interest. Many public health studies also require the use of clustered instead of fully independent data collection designs due to logistical, ethical and cost constraints. For randomized studies, the trials through which clustered data arise are usually called group randomized trials or cluster randomized trials [7, 24]. Such trials have been performed broadly across areas of medicine and public health, most notably in the areas of smoking prevention, physical activity promotion, occupational safety, nutrition, dentistry, and health policy. Clustered designs also arise with the framework of sample surveys, though we do not consider these here. Specific features of clustered data are: the independent sampling unit is the cluster; characteristics of individuals within a cluster tend to be correlated, equally among each other; and the explanatory variable of primary scientific interest, e.g., treatment group, is applied at the cluster level, while data are collected at the individual or within-cluster level.

Many continuous outcomes of interest in public health studies with clustered data have an approximate Gaussian distribution. The analysis of Gaussian clustered data within the framework of a univariate or repeated measures mixed linear model with Gaussian errors is discussed in this dissertation. For simplicity, in this paper, we refer to the explanatory variable of interest as treatment group, though discussion about analyses for difference in treatment

groups applies to any fixed cluster level explanatory variable.

If clustered data are *balanced* so that each cluster contributes the same number of observations, and if data have a common within-cluster correlation and individual error variance across treatment groups, then the set of outcome cluster means are sufficient statistics for inference about treatment group means. That is to say that knowledge of the individual level outcome data gives no additional information about the treatment means over that given by the outcome cluster means. This is true even when the outcome of interest depends on additional covariates other than treatment group, so long as the relationship between the outcome and covariates is the same across treatment groups. With the addition of covariates other than treatment group, knowledge of outcome cluster means and cluster covariate averages suffices for inference about treatment group. Sufficiency of cluster averages for inference about treatment groups is due to the special compound symmetric covariance structure of clustered data.

If data are *unbalanced*, so that a different number of observations is taken in each cluster, then the set of outcome cluster means (with, also, cluster covariate averages, if applicable) are no longer sufficient statistics for inference about treatment group means. This means that inference conducted on cluster means with, potentially, also cluster covariate averages, is different than that conducted on individual level data.

Varnell *et al.* [35] showed that current researchers analyze both cluster means and individual level data. They reviewed group randomized trials published in the *American Journal of Public Health* and *Preventative Medicine* from 1998 to 2002, and showed that of the 47 trials that employed at least one statistical analysis appropriate for group randomized trials (some analyses did not account for the correlation within clusters at all), 15 (32%) analyzed cluster means or another summary statistic and 32 (68%) analyzed individual level data. Analysis of cluster means is often called a "two-stage" model, whereby the cluster means are computed first, often adjusted for covariates through a preliminary model excluding treatment group, and cluster means are the values of the response in a linear model at the second stage [24, p. 112]. Analysis of the individual level data is often called a "one-stage" analysis, where correlated individual level data are analyzed via a mixed effects linear model [24, p. 112].

In this dissertation, we confine interest to Gaussian linear models that include a single fixed effect (e.g., treatment), a random cluster effect, and the usual residual random error.

Such a model is called a two-variance components model [14], two-way nested [27], multi-level [8] or hierarchical model [2], or a one-way mixed model [14]. Though analysis of data from epidemiological studies typically adjust for additional fixed covariates or levels of clustering, it is an important first step to explore the properties of hypothesis tests in the simplest model with no covariates and one level of clustering.

If data have a common within-cluster correlation and common individual error variance, and if data have balanced cluster sizes, both the one-stage and two-stage approaches give the same hypothesis test for the fixed effects [9]. Further, this test is the uniformly most powerful size-α test and has exact null and non-null distributions. Also, this test statistic is derived from closed formed expressions for the maximum likelihood estimates for the fixed effects and variance components, which have known distributions.

If any of the previous conditions about common variance, correlation, or cluster size do not hold, the one-stage and two-stage analysis approaches lead to different tests. No uniformly most powerful size-α test for the fixed effects exist; the unbalanced versions of the test statistics used for balanced data now have only approximate distributions; and closed form expressions for estimates of variance components are no longer available.

Research is needed to study the distributional properties of the hypothesis test statistics for fixed effects in the one-stage and two-stage analysis of unbalanced clustered data. In Chapter 2, enumerations of type I error for hypothesis tests for fixed effects in the two-stage model are presented. Situations of imbalance common to non-randomized clustered data settings are considered there. In Chapter 3, simulations of type I error for hypothesis tests for fixed effects in both the one-stage and two-stage models are presented. These emphasize designs common to group randomized trials. In Chapter 4, a method is described which computes power for the hypothesis test which best controlled type I error in Chapter 3. The aims of these chapters are described in more detail in the next section.

1.2 Aims

1.2.1 Chapter 2, Paper 1

1. Show how to write the null and non-null distributions of the two-stage cluster means model test statistic, with covariance parameters known, as a sum of weighed independent

chi-square random variables. Probabilities for this sum can be computed using Davies [4] algorithm. In the null case, these probabilities are exact; in the alternative case, they are exact given a well-approximated critical value.

2. Produce a SAS/IML module that computes probabilities from the distribution described in Aim 1.

3. Use the modules in Aim 2 to perform an enumeration study showing the bias in type I error for the two-stage cluster means model test statistic for a range of scenarios of imbalance in number of observations per cluster and number of clusters common to non randomized clustered data studies. We will consider the two-stage cluster means model test statistic with cluster means unweighted and weighted by cluster size.

1.2.2 Chapter 3, Paper 2

1. Conduct simulations of type I error for the one-stage individual level and two-stage cluster means model test statistics for a variety of conditions of imbalance common to group randomized trials. Ten tests are considered. For the one-stage model, degrees of freedom will be calculated (1-2) by the method of Kenward and Roger [13] and (3-4) as the number of clusters minus the number of treatment groups. Both methods will include simulations with variance components unconstrained and constrained to be positive. For the two-stage model, the following weight matrices will be considered: (5) unweighted, (6) weighted by cluster size, (7) weighted by inverse of cluster size, (8) weighted by the inverse of the sample variance of each cluster mean, and (9-10) weighted by the inverse of the theoretical variance of each cluster mean, with variance components estimated from the entire data. The last two stage analysis will be performed with variance components unconstrained and constrained to be positive.

2. Suggest which of the tests in Aim 1 controls type I error for the most scenarios of imbalance.

3. Suggest conditions of imbalance under which each test as well as more than one test provides an unbiased test for the fixed effects.

1.2.3 Chapter 4, Paper 3

1. Show how to compute power for unbalanced clustered data in the two stage analysis of cluster means with means weighted by the inverse of estimates of their theoretical variance.

2. Illustrate use of this method using data from a study on adolescent drinking behavior.

In the remainder of this chapter, I discuss literature relevant to each of these topics; material for each paper is developed in more detail in Chapters 2, 3, and 4. Chapter 2, 3, and 4 were written as stand alone papers, and so repeat some of the literature review presented in this chapter.

1.3 Notation

1.3.1 Matrix Notation

This section describes the notational conventions used in this document. Lower case bold indicates a (column) vector, upper case bold a matrix. Upper case italics indicates a non-matrix random variable. Matrix notation dominates over random variable notation, so that randomness of a matrix must be inferred from context.

We follow the notation of McCulloch and Searle [19], Appendix M, Section 3, to conveniently denote stacked column vectors and diagonal matrices with similar notation. Define the indices i and j such that $i = 1, ..., a$ and $j = 1, ..., b$. Let $\boldsymbol{u} = \left\{ {}_c\, \boldsymbol{u}_{ij} \right\}$ denote the stacked column vector $\boldsymbol{u}$, where $\boldsymbol{u} = \left\{ \boldsymbol{u}'_{11} \quad \boldsymbol{u}'_{12} \quad \dots \quad \boldsymbol{u}'_{ij} \quad \dots \quad \boldsymbol{u}'_{ab} \right\}'$. Let $\boldsymbol{U} = \left\{ {}_d\, \boldsymbol{U}_{ij} \right\}$ denote the diagonal matrix $\boldsymbol{U}$ with diagonal elements $\boldsymbol{U}_{11}, \boldsymbol{U}_{12}, ..., \boldsymbol{U}_{ij}, ..., \boldsymbol{U}_{ab}$. These notations are identical except for the subscript on the opening brace; the subscript c denotes a stacked column vector and the subscript d denotes a diagonal matrix.

Kronecker product multiplication of matrix $\boldsymbol{A}$ by matrix $\boldsymbol{B}$ is denoted by $\boldsymbol{A} \otimes \boldsymbol{B}$,where

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & ... & a_{1a}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & ... & a_{2a}\boldsymbol{B} \\ ... & ... & ... & .. \\ a_{b1}\boldsymbol{B} & a_{b2}\boldsymbol{B} & ... & a_{ab}\boldsymbol{B} \end{bmatrix},$$

and a_{ij} is the element in the i-th row and j-th column of matrix $\boldsymbol{A}$. All other matrix operators are defined as in standard practice; Muller and Stewart [23] or Schott [30] give details.

1.3.2 Distributions

Let $\boldsymbol{x} \sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicate that the vector $\boldsymbol{x}$ $(N \times 1)$ follows an N-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $X \sim \mathcal{F}(\nu_1, \nu_2, \omega)$ indicate that the random variable X has a noncentral F distribution with ν_1 numerator degrees of freedom, ν_2 denominator degrees of freedom, and noncentrality ω. Let $X \sim \chi^2(\nu_1, \omega)$ indicate that X has a non-central chi-square distribution with ν_1 degrees of freedom and noncentrality ω. With zero noncentralities, both the noncentral F and χ^2 distributions reduce to central versions. Kotz *et al.* [17] gives detailed information about these distributions.

1.3.3 Data Indices and Model Notation

Theory and results in this document are discussed within the framework of the general linear mixed model and the general linear univariate model. Notational conventions from these areas are used heavily in this document, e.g., Verbeke and Molenberghs [36] or Muller and Stewart [23]. Such notation can differ from other notational schemes that also would have been defensible, namely, that used in any of multivariate, hierarchical, or multi-level linear models or in the field of group or cluster randomized trials. When necessary for clarity, notation from these fields must be employed. In particular, because this dissertation focuses heavily on properties of balanced versus unbalanced data, we make different choices of notation for total number of observations, number of independent sampling units, and number of observations per cluster, than those usually made in traditional mixed model notation. Table 1.1 summarizes these notational differences. Table 1.2 summarizes the notation used in this document to describe the structure of clustered data.

Table 1.1: Notation in This Document Compared to Traditional Mixed Model Notation

	This Document	Mixed Model
Total number of observations	N	n
Number of independent sampling units (clusters)	m	N
Number of observations in the i-th cluster of the h-th treatment group when cluster sizes are balanced	n	n_{hi}
Number of observations in the i-th cluster of the h-th treatment group when cluster sizes are unbalanced	n_{hi}	n_{hi}

Table 1.2: Summary of Non Matrix Notation

Symbol	**Definition**
	Indicies
$h = 1, \ldots, g$	Indexes treatment groups
$i = 1, \ldots, m_h$	Indexes clusters within treatment group
$j = 1, \ldots, n_{hi}$	Indexes observations within cluster
	Numbers of Clusters and Observations
g	Number of treatment groups
m_h	Number of clusters in treatment group h
$m = \sum_{h=1}^{g} m_h$	Total number of clusters
n_{hi}	Number of observations within a cluster when cluster sizes are unequal
n	Number of observations within a cluster when cluster sizes are equal
$n_h = \sum_{i=1}^{m_h} n_{hi}$	Number of observations in treatment group h
$N = \sum_{h=1}^{g} \sum_{i=1}^{m_h} n_{hi}$	Total number of observations
	Outcome Notation
y_{hij}	Outcome for observation j of cluster i in treatment group h
$\overline{y}_{hi} = \frac{1}{n_{hi}} \sum_{j=1}^{n_{hi}} y_{hij}$	Outcome mean for cluster i in treatment group h
$\overline{y}_h = \frac{1}{m_h} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} y_{hij}$	Outcome mean for treatment group h

1.4 Statement of Models and Hypothesis

1.4.1 One-Stage Model

Define a linear model for continuous Gaussian outcome $\boldsymbol{y}_1$ that includes fixed effects given in $\boldsymbol{\beta}$ $(g \times 1)$, a random effect for cluster given in $\boldsymbol{b}$ $(m \times 1)$, and a random error, $\boldsymbol{e}_1$ $(N \times 1)$:

$$\boldsymbol{y}_1 = \boldsymbol{X}_1\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{b} + \boldsymbol{e}_1. \tag{1.1}$$

The matrices $\boldsymbol{X}_1$ $(N \times g)$ and $\boldsymbol{Z}_1$ $(N \times m)$ are design matrices for the fixed and random effects, respectively. This review assumes $\boldsymbol{X}_1$ contains only an effect for treatment group and $\boldsymbol{Z}_1$ only an effect for cluster.

Vectors or matrices $\boldsymbol{y}_1$, $\boldsymbol{X}_1$, $\boldsymbol{Z}_1$, and $\boldsymbol{e}_1$ are stacked by treatment group and cluster so that $\boldsymbol{y}_1 = \{_c\, \boldsymbol{y}_{1,hi}\}$, $\boldsymbol{X}_1 = \{_c\, \boldsymbol{X}_{1,hi}\}$, $\boldsymbol{Z}_1 = \{_c\, \boldsymbol{Z}_{1,hi}\}$, and $\boldsymbol{e}_1 = \{_c\, \boldsymbol{e}_{1,hi}\}$. Without loss of generality, assume the fixed effects design matrix $\boldsymbol{X}_1$ has a cell mean coding for treatment group so that $\boldsymbol{X}_1 = \{_d\, \mathbf{1}_{n_h}\}$. The design matrix for the random cluster effect is $\boldsymbol{Z}_1 = \{_d\, \mathbf{1}_{n_{hi}}\}$. When data have balanced cluster sizes, these simplify to $\boldsymbol{X}_1 = \{_d\, \mathbf{1}_{m_h n}\}$ and $\boldsymbol{Z}_1 = \boldsymbol{I}_m \otimes \mathbf{1}_n$.

We assume $\boldsymbol{b} \sim \mathcal{N}_m\left(\mathbf{0}, \sigma_c^2 \boldsymbol{I}_m\right)$ independently of $\boldsymbol{e}_1 \sim \mathcal{N}_N\left(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_N\right)$ so that:

$$\boldsymbol{y}_1 \sim \mathcal{N}_N\left(\boldsymbol{X}_1\boldsymbol{\beta}, \boldsymbol{\Sigma}_1\right),$$

where the covariance matrix $\boldsymbol{\Sigma}_1$ $(N \times N)$ is compound symmetric and has the form:

$$\boldsymbol{\Sigma}_1 = \sigma_c^2 \boldsymbol{Z}_1\boldsymbol{Z}_1' + \sigma_e^2 \boldsymbol{I}_N = \{_d\, \sigma_c^2 \mathbf{1}_{n_{hi}}\mathbf{1}_{n_{hi}}' + \sigma_e^2 \boldsymbol{I}_{n_{hi}}\}.$$

$\boldsymbol{\Sigma}_1$ may be expressed in terms of the total variance, σ_y^2, and within cluster correlation, ρ, as:

$$\boldsymbol{\Sigma}_1 = \sigma_y^2\{_d\, \mathbf{1}_{n_{hi}}\mathbf{1}_{n_{hi}}'\rho + \boldsymbol{I}_{n_{hi}}(1-\rho)\},$$

where $\rho = \sigma_c^2/\left(\sigma_c^2 + \sigma_e^2\right)$ and $\sigma_y^2 = \sigma_c^2 + \sigma_e^2$ or, equivalently, $\sigma_c^2 = \sigma_y^2\rho$ and $\sigma_e^2 = \sigma_y^2(1-\rho)$. Implicit in construction of $\boldsymbol{\Sigma}_1$ is the assumption that data across all treatment groups have the same variance parameters. When data have balanced cluster sizes the covariance matrix $\boldsymbol{\Sigma}_1$ simplifies to:

$$\boldsymbol{\Sigma}_1 = \boldsymbol{I}_m \otimes \left(\sigma_c^2 \mathbf{1}_n\mathbf{1}_n' + \sigma_e^2 \boldsymbol{I}_n\right) = \boldsymbol{I}_m \otimes \sigma_y^2\{_d\, \mathbf{1}_n\mathbf{1}_n'\rho + \boldsymbol{I}_n(1-\rho)\}.$$

1.4.2 Two-Stage Model

To transform from a model for individual level data to a model for cluster means, pre-multiply model (1.1) by the matrix $\boldsymbol{T}_1$ $(m \times N)$, where $\boldsymbol{T}_1 = \left\{ {}_d\, \mathbf{1}'_{n_{hi}}/n_{hi} \right\}$. This yields a model for $\boldsymbol{y}_2$ $(m \times 1) = \boldsymbol{T}_1 \boldsymbol{y}_1$ where:

$$\boldsymbol{y}_2 = \boldsymbol{X}_2 \boldsymbol{\beta} + \boldsymbol{Z}_2 \boldsymbol{b} + \boldsymbol{e}_2, \tag{1.2}$$

and $\boldsymbol{X}_2$ $(m \times g) = \boldsymbol{T}_1 \boldsymbol{X}_1$, $\boldsymbol{Z}_2$ $(m \times m) = \boldsymbol{T}_1 \boldsymbol{Z}_1$, and $\boldsymbol{e}_2$ $(m \times 1) = \boldsymbol{T}_1 \boldsymbol{e}_1$. Parameters in $\boldsymbol{\beta}$ and $\boldsymbol{b}$ were not affected by the transformation.

The vector of outcomes, $\boldsymbol{y}_2$, and of random errors, $\boldsymbol{e}_2$, contain cluster averages, so that $\boldsymbol{y}_2 = \left\{ {}_c\, \overline{y}_{hi} \right\}$ and $\boldsymbol{e}_2 = \left\{ {}_c\, \overline{e}_{hi} \right\}$. The fixed and random effects design matrices are $\boldsymbol{X}_2 = \left\{ {}_d\, \mathbf{1}'_{n_{hi}}/n_{hi} \right\} \left\{ {}_d\, \mathbf{1}_{n_h} \right\} = \left\{ {}_d\, \mathbf{1}_{m_h} \right\}$ and $\boldsymbol{Z}_2 = \left\{ {}_d\, \mathbf{1}'_{n_{hi}}/n_{hi} \right\} \left\{ {}_d\, \mathbf{1}_{n_{hi}} \right\} = \boldsymbol{I}_m$.

In line with previous assumptions, we assume $\boldsymbol{b} \sim \mathcal{N}_m \left(\mathbf{0}, \sigma_c^2 \boldsymbol{I}_m \right)$ independently of $\boldsymbol{e}_2 \sim \mathcal{N}_m \left(\mathbf{0}, \sigma_e^2 \boldsymbol{T}_1 \boldsymbol{T}'_1 \right)$ so that:

$$\boldsymbol{y}_2 \sim \mathcal{N}_m \left(\boldsymbol{X}_2 \boldsymbol{\beta}, \boldsymbol{\Sigma}_2 \right),$$

where $\boldsymbol{\Sigma}_2$ $(m \times m)$ is given by:

$$\boldsymbol{\Sigma}_2 = \boldsymbol{T}_1 \boldsymbol{\Sigma}_1 \boldsymbol{T}'_1 = \left\{ {}_d\, \sigma_c^2 + \sigma_e^2/n_{hi} \right\}.$$

In terms of the alternate parameterization with $\left(\sigma_y^2, \rho \right)$ instead of $\left(\sigma_e^2, \sigma_c^2 \right)$:

$$\boldsymbol{\Sigma}_2 = \sigma_y^2 \left\{ {}_d\, \left[1 + (n_{hi} - 1)\, \rho \right] / n_{hi} \right\}.$$

1.4.3 General Linear Hypothesis

Define a vector of secondary contrast parameters, $\boldsymbol{\theta}$ $(a \times 1) = \boldsymbol{C}\boldsymbol{\beta}$, where $\boldsymbol{C}$ $(a \times g)$ contains desired contrasts for the fixed effects. For clustered data $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ in models 1.1 and 1.2, elements of $\boldsymbol{\theta}$ are linear combinations of cluster means. We study the two-sided general linear hypothesis (GLH):

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \text{ versus } H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{1.3}$$

In most hypotheses of interest, $\boldsymbol{\theta}_0 = \mathbf{0}$. Such a hypothesis test describes differences in the fixed effects only. We do not consider tests for variance components or ratios of variance components in this dissertation. If $\boldsymbol{\theta}$ is estimable, requiring $\boldsymbol{C}$ to have full row rank [rank$(\boldsymbol{C}) = a$] ensures the GLH is testable. $\boldsymbol{\theta}$ is estimable if and only if $\boldsymbol{C} = \boldsymbol{C} \left(\boldsymbol{X}'\boldsymbol{X} \right)^- \left(\boldsymbol{X}'\boldsymbol{X} \right)$. Note that this

requirement is automatically satisfied if $\boldsymbol{X}$ is full rank, so that $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists [21].

1.5 Literature Review

1.5.1 Hypothesis Testing for Clustered Data with Balanced Cluster Sizes

The analysis of clustered data has special properties when data have balanced cluster sizes. This section describes estimation and hypothesis testing of fixed effects for such balanced data. In practice, these methods are applied to unbalanced clustered data as well. Section 1.5.2 describes the properties of methods of estimation and hypothesis testing for balanced clustered data when applied to unbalanced data.

1.5.1.1 Estimation of Fixed Effects for the One Stage Model for Individual Data

Consider the one-stage model for individual level data $\boldsymbol{y}_1 \sim \mathcal{N}_N(\boldsymbol{X}_1\boldsymbol{\beta}, \boldsymbol{\Sigma}_1)$ given in model 1.1. When $\boldsymbol{\Sigma}_1$ is unknown, and therefore contains nuisance parameters which must be estimated, the restricted maximum likelihood (REML) estimator for $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{y}_1,$$

where $\widehat{\boldsymbol{\Sigma}}_1\,(N \times N)$ is the REML estimator for $\boldsymbol{\Sigma}_1$. When individual level clustered data $\boldsymbol{y}_1$ have balanced cluster sizes, that is, when $n_{hi} \equiv n$ for all h,i, the estimator $\widehat{\boldsymbol{\Sigma}}_1$ can be stated in terms of a Kronecker product of the same covariance matrix for all clusters. That is, $\widehat{\boldsymbol{\Sigma}}_1 = \boldsymbol{I}_m \otimes \left\{\widehat{\sigma}_y^2\left[\mathbf{1}_n\mathbf{1}_n'\widehat{\rho} + \boldsymbol{I}_n(1-\widehat{\rho})\right]\right\}$. Because of this, the inverse of $\widehat{\boldsymbol{\Sigma}}_1$ can be written with the closed form expression:

$$\widehat{\boldsymbol{\Sigma}}_1^{-1} = \boldsymbol{I}_m \otimes \frac{1}{\widehat{\sigma}_y^2(1-\widehat{\rho})}\left\{\boldsymbol{I}_n - \frac{\widehat{\rho}}{[1+(n-1)\widehat{\rho}]}\mathbf{1}_n\mathbf{1}_n'\right\}.$$

For balanced data with no covariates, $\boldsymbol{X}_1 = \left\{{}_d\,\mathbf{1}_{m_h n}\right\}$, and we can show that $\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1} = \left\{\widehat{\sigma}_y^2[1+(n-1)\widehat{\rho}]\right\}^{-1}\boldsymbol{X}_1'$. Thus, for balanced data:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{y}_1 \\
&= \left(\left\{\widehat{\sigma}_y^2[1+(n-1)\widehat{\rho}]\right\}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\left\{\widehat{\sigma}_y^2[1+(n-1)\widehat{\rho}]\right\}^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1 \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1.
\end{aligned}$$

We now have an estimator for the fixed effects that does not depend on the variance components. That is, for balanced data, the weighted least squares and ordinary least squares estimators of

$\boldsymbol{\beta}$ coincide. Puntanen and Styan [26] and Tian and Wiens [34] gave a comprehensive review of when weighted or ordinary least squares estimators coincide for general data $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1.5.1.2 Estimation of Fixed Effects for the Two Stage Model for Cluster Means

Consider the two-stage model for cluster means $\boldsymbol{y}_2 \sim \mathcal{N}_m(\boldsymbol{X}_2\boldsymbol{\beta}, \boldsymbol{\Sigma}_2)$ given in model 1.2. When data are balanced, $n_{hi} \equiv n$ for all h, i, so that the cluster means have covariance:

$$\boldsymbol{\Sigma}_2 = \boldsymbol{I}_m \otimes \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}.$$

That is, for balanced data, all cluster means have the same variance, and $\boldsymbol{\Sigma}_2$ can be written as $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{I}_m$ where: $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$. Independence, normality, and homogeneity of errors of the cluster means meet the assumptions of the general linear univariate model (GLUM).

In the general linear univariate model, the best linear unbiased and maximum likelihood estimator for the fixed effects, $\boldsymbol{\beta}$, is:

$$\widehat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{y}_2.$$

1.5.1.3 Equivalence of Estimators from One and Two Stage Models

Matrix algebra allows showing that:

$$\boldsymbol{X}_1'\boldsymbol{X}_1 = \left\{{}_d\,\mathbf{1}_{m_h n}'\right\}\left\{{}_d\,\mathbf{1}_{m_h n}\right\} = \left\{{}_d\,n m_h\right\}$$

$$\boldsymbol{X}_1'\boldsymbol{y}_1 = \left\{{}_d\,\mathbf{1}_{m_h n}'\right\}\left\{{}_c\,\boldsymbol{y}_{1,hi}\right\} = \left\{{}_c \sum_{i=1}^{m_h} n\overline{y}_{hi}\right\}$$

as well as:

$$\boldsymbol{X}_2'\boldsymbol{X}_2 = \left\{{}_d\,\mathbf{1}_{m_h}'\right\}\left\{{}_d\,\mathbf{1}_{m_h}\right\} = \left\{{}_d\,m_h\right\}$$

$$\boldsymbol{X}_2\boldsymbol{y}_2 = \left(\mathbf{1}_{m_h}'\right)\left\{{}_c\,\overline{y}_{hi}\right\} = \left\{{}_c \sum_{i=1}^{m_h} \overline{y}_{hi}\right\}.$$

Thus:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1 = \left\{{}_d\,n m_h\right\}\left\{{}_c \sum_{i=1}^{m_h} n\overline{y}_{hi}\right\} = \left\{{}_c\,\overline{y}_h\right\}$$

and:

$$\widehat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{y}_2 = \left\{{}_d\,m_h\right\}\left\{{}_c \sum_{i=1}^{m_h} \overline{y}_{hi}\right\} = \left\{{}_c\,\overline{y}_h\right\}.$$

That is, for data with balanced cluster sizes, the population treatment means in $\boldsymbol{\beta}$ are estimated by the sample treatment means in both the one-stage and two-stage models.

1.5.1.4 Hypothesis Test for Fixed Effects

The equivalent estimators given in sections 1.5.1.1 and 1.5.1.2 may be written as

$$\widehat{\boldsymbol{\beta}}_s = \left(\boldsymbol{X}'_s\boldsymbol{X}_s\right)^{-1}\boldsymbol{X}'_s\boldsymbol{y}_s$$

where $s = 1, 2$. Using theory of quadratic forms in normal vectors, $\widehat{\boldsymbol{\beta}}_s \sim \mathcal{N}_g\left[\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}'_s\boldsymbol{X}_s\right)^{-1}\right]$, where $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$. Further, an estimator for desired contrasts, $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$, may be estimated with $\widehat{\boldsymbol{\theta}}_s = \boldsymbol{C}\widehat{\boldsymbol{\beta}}_s$, which has distribution $\widehat{\boldsymbol{\theta}}_s \sim \mathcal{N}_a\left[\boldsymbol{\theta}, \sigma^2\boldsymbol{C}\left(\boldsymbol{X}'_s\boldsymbol{X}_s\right)^{-1}\boldsymbol{C}'\right]$.

Using theory for a general linear univariate model, a uniformly most powerful test size α for the GLH is given by:

$$T_s = \left(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_0\right)'\left[\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_s\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_0\right)/a,$$

where $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_s\right)$ is the estimated variance of $\widehat{\boldsymbol{\theta}}_s$, that is $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_s\right) = \widehat{\sigma}^2\left[\boldsymbol{C}\left(\boldsymbol{X}'_s\boldsymbol{X}_s\right)^{-1}\boldsymbol{C}'\right]$. The quantity $\widehat{\sigma}^2$ denotes the restricted maximum likelihood estimator for σ^2, discussed in the next section. This test statistic can be shown to have distribution:

$$T_s \sim \mathcal{F}\left(a, m-g, \omega\right),$$

where $\omega = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)'\left[\boldsymbol{C}\left(\boldsymbol{X}'_s\boldsymbol{X}_s\right)^{-1}\boldsymbol{C}'\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)/\sigma^2$.

1.5.1.5 Estimation of Variance Components

In the two stage analysis of cluster means, variance components $\left(\sigma_y^2, \rho\right)$ or $\left(\sigma_c^2, \sigma_e^2\right)$ are not separately estimable so that the linear combination $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$ is estimated. An estimator for σ^2 in the two stage analysis of cluster means is:

$$\widehat{\sigma}^2 = \boldsymbol{y}'_2\left[\boldsymbol{I}_m - \boldsymbol{X}_2\left(\boldsymbol{X}'_2\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}'_2\right]\boldsymbol{y}_2/\left(m-g\right).$$

This estimator can be derived as an restricted maximum likelihood and ANOVA estimator $\widehat{\sigma}^2 = \mathrm{SSE}_2/(m-g)$, where SSE_2 denotes the sums of squares error.

In the one stage analysis of individual level data, though an estimator of the linear combination $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\} = \sigma_c^2 + \sigma_e^2/n$ is needed, current statistical software, designed for the estimation of parameters of general covariance structures, estimates the variance components separately in the parameterization $\left(\sigma_c^2, \sigma_e^2\right)$.

Define the sums of squares due to cluster and error, respectively, as:

$$\mathrm{SSC}_1 = \boldsymbol{y}_1' \left[\boldsymbol{Z}_1 \left(\boldsymbol{Z}_1' \boldsymbol{Z}_1 \right)^{-1} \boldsymbol{Z}_1 - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1 \right] \boldsymbol{y}_1$$

$$\mathrm{SSE}_1 = \boldsymbol{y}_1' \left[\boldsymbol{I}_N - \boldsymbol{Z}_1 \left(\boldsymbol{Z}_1' \boldsymbol{Z}_1 \right)^{-1} \boldsymbol{Z}_1' \right] \boldsymbol{y}_1.$$

as well as mean squares due to cluster and error, respectively:

$$\mathrm{MSC}_1 = \mathrm{SSC}_1 / \left(m - g \right)$$

$$\mathrm{MSE}_1 = \mathrm{SSE}_1 / \left(N - m \right).$$

The parameterization $\left(\sigma_c^2, \sigma_e^2 \right)$ requires estimates of both σ_c^2 and σ_c^2 to be positive, since σ_c^2 and σ_e^2 are defined as variances. As such, several sources, e.g. Searle [31, p .419], have pointed out that restricted maximum likelihood estimators for σ_c^2 and σ_e^2 are given by:

$$\widehat{\sigma}_c^2 = \left(\mathrm{MSC}_1 - \mathrm{MSE}_1 \right) / n$$

$$\widehat{\sigma}_e^2 = \mathrm{MSE}_1$$

when $\mathrm{MSC}_1 \geq \mathrm{MSE}_1$ (that is, when $\widehat{\sigma}_c^2$ is positive) and

$$\widehat{\sigma}_c^2 = 0$$

$$\widehat{\sigma}_e^2 = \mathrm{SST}_1 / \left(N - m \right)$$

when $\mathrm{MSC}_1 < \mathrm{MSE}_1$, where $\mathrm{SST}_1 = \mathrm{SSC}_1 + \mathrm{SSE}_1$ denotes the total sums of squares of the individual level data. The probability that $\widehat{\sigma}_c^2 < 0$ is:

$$\Pr \left\{ \widehat{\sigma}_c^2 < 0 \right\} = \Pr \left\{ \mathcal{F}_{m-1, N-m} < 1 / \left[1 + n\rho / \left(1 - \rho \right) \right] \right\}.$$

It can be shown that when $\mathrm{MSC}_1 \geq \mathrm{MSE}_1$, the estimator $\widehat{\sigma}^2 = \widehat{\sigma}_c^2 + \widehat{\sigma}_e^2 / n$ is equivalent to the best linear unbiased and maximum likelihood estimator obtained in the two-stage analysis; however this is not the case when $\mathrm{MSC}_1 < \mathrm{MSE}_1$. That is, when $\mathrm{MSC}_1 < \mathrm{MSE}_1$, the linear combination of restricted maximum likelihood estimators for each of σ_c^2 and σ_e^2 is NOT the restricted maximum likelihood estimator for the linear combination σ^2. The restricted maximum likelihood estimator for σ^2 is obtained only when variance components estimators are $\widehat{\sigma}_c^2 = \left(\mathrm{MSC}_1 - \mathrm{MSE}_1 \right) / n$ and $\widehat{\sigma}_e^2 = \mathrm{MSE}_1$, and the estimator $\widehat{\sigma}_c^2$ is allowed to be negative.

Default behavior of SAS PROC MIXED is to constrain estimates of variance components to be positive; simulations in this paper explore the ramifications of this choice.

1.5.2 Hypothesis Testing for Clustered Data with Unbalanced Cluster Sizes

1.5.3 One Stage Model for Individual Data

When cluster sizes are unbalanced, the weighted least squares estimator $\widehat{\boldsymbol{\beta}}_1$, given in section 1.5.1.1 as:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{y}_1,$$

is no longer equivalent to an ordinary least squares estimator. That is, estimation of fixed effects now requires estimation of the variance components.

Recall that the structure of $\boldsymbol{\Sigma}_1$ for unbalanced data is:

$$\boldsymbol{\Sigma}_1 = \left\{{}_d\,\sigma_c^2\mathbf{1}_{n_{hi}}\mathbf{1}_{n_{hi}}' + \sigma_e^2\boldsymbol{I}_{n_{hi}}\right\} = \sigma_y^2\left\{{}_d\,\mathbf{1}_{n_{hi}}\mathbf{1}_{n_{hi}}'\rho + \boldsymbol{I}_{n_{hi}}(1-\rho)\right\}.$$

When data are unbalanced, no closed form expressions exist for estimates of the variance components in either parameterization; estimates must be obtained by an iterative procedure such as Newton-Raphson iteration or the EM algorithm [5].

Construction of a hypothesis test for $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$ requires knowledge of the distribution of $\widehat{\boldsymbol{\beta}}_1$. The estimator $\widehat{\boldsymbol{\beta}}_1$ is unbiased, so that $\mathcal{E}\left(\widehat{\boldsymbol{\beta}}_1\right) = \boldsymbol{\beta}$; however, no closed form expression exists for its variance. The common strategy is to approximately estimate this as $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\beta}}_1\right) \approx \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}$. Kacker and Harville [12] and Dempster *et al.* [6] pointed out that this underestimates the true variability in $\widehat{\boldsymbol{\beta}}_1$, since $\left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}$ is an estimate of the variance of $\widetilde{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{y}_1$ not of the variance of $\widehat{\boldsymbol{\beta}}_1$.

As in hypothesis testing with balanced data, a hypothesis test for the general linear hypothesis is given by the Wald style test statistic:

$$T_1 = \left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\right)'\left[\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_1\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\right)/a.$$

Since closed form expressions do not exist for the elements of $\widehat{\boldsymbol{\Sigma}}_1$, T_1 cannot be written explicitly as a quadratic form, and thus its exact distribution is unknown. Applying large sample theory gives $T_1 \xrightarrow{D} \chi^2(a, \omega)$. Given a random variable X with $X \sim \mathcal{F}(\nu_1, \nu_2, \omega)$, as $\nu_2 \longrightarrow \infty$, $X \xrightarrow{D} Y$ where $Y \sim \chi^2(\nu_1, \omega)$. For this reason, T_1 is usually given an approximate $F(a, \nu_2, \omega)$ distribution in order to combat the underestimate of the variance of $\widehat{\boldsymbol{\beta}}_1$. Several methods exist to estimate the denominator degrees of freedom of T_1. In most cases, none can be shown to be uniformly superior [18].

One such method is that proposed by Kenward and Roger [13], who multiply T_1 by an inflation factor to account for the additional variability in $\mathcal{V}\left(\widehat{\boldsymbol{\beta}}_1\right)$ introduced by estimating $\boldsymbol{\Sigma}_1$. Satterthwaite [28] style degrees of freedom are then computed for this inflated statistic. This approximation has not been studied thoroughly in small clustered data settings, and has been shown to be biased in settings with other types of small sample data [3, 15, 29]. Another common choice for denominator degrees of freedom is that for the analysis of balanced data, $\text{ddf} = m - g$.

1.5.4 Two Stage Model for Cluster Means

From section 1.5.1.2, cluster means with balanced cluster sizes may be analyzed via a general linear univariate model. The general linear univariate model assumes $\boldsymbol{\Sigma}_2$ is proportional to an identity matrix, that is $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{I}$, where as before $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$.

An optimal test for the general linear hypothesis can be derived with the less restrictive assumption that $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{W}^{-1}$, that is, that the covariance matrix is known up to a constant weight matrix $\boldsymbol{W}$ [19]. In this case, the best linear unbiased and maximum likelihood estimator for the fixed effects is:

$$\widehat{\boldsymbol{\beta}}_{2w} = \left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{y}_2.$$

With $\widehat{\boldsymbol{\beta}}_{2w} \sim \mathcal{N}_g\left[\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\right]$ and thus, $\widehat{\boldsymbol{\theta}}_{2w} = \boldsymbol{C}\widehat{\boldsymbol{\beta}}_{2w} \sim \mathcal{N}_a\left[\boldsymbol{\theta}, \sigma^2\boldsymbol{C}\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{C}'\right]$, a uniformly most powerful size-α test for the fixed effects can be shown to be given by:

$$T_{2w} = \left(\widehat{\boldsymbol{\theta}}_{2w} - \boldsymbol{\theta}_0\right)'\left[\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_{2w}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}_{2w} - \boldsymbol{\theta}_0\right)/a,$$

where $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_{2w}\right) = \widehat{\sigma}_w^2\boldsymbol{C}\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{C}'$. The restricted maximum likelihood estimator for σ^2 is:

$$\widehat{\sigma}_w^2 = \boldsymbol{y}_2'\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{W}\right]\boldsymbol{y}_2/\left(m-g\right).$$

The statistic T_{2w} has distribution $T_{2w} \sim \mathcal{F}\left(\nu_1,\, \nu_2,\, \omega_w\right)$, where the noncentrality in the weighted model is $\omega_w = \left(\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{\theta}_0\right)'\left[\boldsymbol{C}\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{C}'\right]^{-1}\left(\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{\theta}_0\right)/\sigma^2$.

The exact distribution of the test statistic T_w depends on the assumption that $\boldsymbol{\Sigma}_2$ can be written in the form $\boldsymbol{\Sigma}_2 = \sigma^2\boldsymbol{W}^{-1}$ where $\boldsymbol{W}$ is known. When this doesn't hold T_{2w} has only an approximate $\mathcal{F}$ distribution. Cluster means derived from unbalanced clustered data with $\boldsymbol{\Sigma}_2 = \left(\sigma_y^2\right)\left\{{}_d\left[1 + (n_{hi} - 1)\,\rho\right]/n_{hi}\right\}$ do not have this form since variance components must be

estimated; however, many types of weights have $\boldsymbol{W}$ such that $\boldsymbol{\Sigma}_2 \approx \sigma^2 \boldsymbol{W}^{-1}$.

If ρ is small and cluster sizes are small, $\boldsymbol{\Sigma}_2 \approx \left(\sigma_y^2\right) \left\{_d\, 1/n_{hi}\right\}$, so that weighting by cluster sizes is an appropriate choice. Forgetting to invert the weights also may lead researchers to erroneously weight by the inverse of cluster size. If cluster sizes are not highly variable, $\boldsymbol{\Sigma}_2 \approx \sigma^2 \boldsymbol{W}^{-1}$ directly, so an analysis of unweighted means is performed.

In an attempt to estimate the variance components in the weights, data analysts also choose $\boldsymbol{W} = \left\{_d [n_{hi}\,(n_{hi}-1)] \,/\, [\boldsymbol{y}'_{hi}\boldsymbol{y}_{hi} - n_{hi}\bar{y}_{1,hi}]\right\}$, the diagonal matrix of inverses of the estimated sample variance of each cluster mean, or $\boldsymbol{W} = \left[\left(\widehat{\sigma}_y^2\right) \left\{_d [1 + (n_{hi}-1)\,\widehat{\rho}\,] / n_{hi}\right\}\right]^{-1}$, the diagonal matrix of the inverse of estimates of the theoretical variance of each cluster mean with variance components estimated from all the data. Such estimates of variance components from the data can be constrained to be positive or allowed to be negative.

1.6 Motivating Data Example

The Trial of Activity in Adolescent Girls (TAAG) [32] was designed to evaluate an intervention to reduce the age-related decline in moderate to vigorous physical activity (MVPA) in middle school girls. The primary endpoint is the mean difference in intensity-weighted minutes (MET minutes) of MVPA between intervention and control schools. 18 schools were randomized to each of intervention and control for a total of 36 schools. Baseline MVPA data were collected on a sample of 60 6th grade girls per school, then again two years later on a sample of 120 8th graders per school. Students were sampled at both time points so that data were not necessarily collected on the same girls at both data collection points. The primary planned analysis was a two-stage analysis of endpoint MVPA school means adjusted for baseline MVPA school means. Further details can be found in Stevens *et al.* [32] and Murray *et al.* [25].

One-stage analysis methods on individual level data which properly adjust for baseline values cannot be conducted on the TAAG data, because the independent cross-sectional sampling design resulted in incomplete overlap of girls at the two time points. Nonetheless, one can analyze these data via both two-stage and one-stage methods if baseline data is omitted. We perform such an analysis below for instructional purposes only; this analysis was not one actually conducted for this trial.

Suppose the study had pre-planned subgroup analyses by race. The majority of girls in

the study belonged to three ethnic groups: non-Hispanic white and black, and Hispanic (of any race). All race groups were not equally represented in all schools. To estimate the treatment effect among black students, we considered only those schools with at least 2 African American students, leading to inclusion of 12 intervention schools and 13 control schools. Because the racial profiles of each school vary widely, each cluster (school) will contribute different numbers of observations to these data. Clusters sizes in the intervention and control schools were {2, 2, 6, 12, 18, 29, 30, 31, 36, 52, 60, 81}, (mean = 29.9), and {5, 7, 7, 11, 13, 14, 14, 18, 23, 28, 35, 62, 69}, (mean = 23.5), respectively. These cluster sizes are naturally much more variable that those in the planned primary analysis, but illustrate how extremely unbalanced data can easily be obtained in an analysis of a subgroup of well-balanced data.

Table 1.3 gives p-values for a test for difference in average MVPA for intervention versus control schools within the framework of each of the following models :

1. One-stage analysis via a mixed effects model with denominator degrees of freedom calculated by the method of Kenward and Roger [13].

2. One-stage analysis via a mixed effects model with denominator degrees of freedom equal to $m - g$.

3. Two-stage analysis of unweighted cluster means

4. Two-stage analysis of cluster means weighted by cluster size

5. Two-stage analysis of cluster means weighted by the inverse of cluster size

6. Two-stage analysis of cluster means weighted by the inverses of sample variances of each cluster mean

7. Two-stage analysis of cluster means weighted by the inverse of the theoretical variance, with variance components estimated with the entire data.

Estimates of variance components for these data were positive.

One stage strategies 1 and 2 and two-stage strategy 7 compute similar, though not equivalent p-values. For these data with unequal number of clusters per intervention group and maximum

cluster size equal to 15 or 40 times the minimum cluster size, two stage approaches 3 - 6 give widely different results.

Table 1.3: Type I Error for One Stage and Two Stage Analyses

Analysis	Estimated MVPA Mean (SE) Intervention Group	Estimated MVPA Mean (SE) Control Group	Estimated Difference	F	Denom DF	P-value
1	20.35 (0.90)	18.85 (0.90)	1.50 (1.27)	1.18	15.6	0.26
2	20.35 (0.89)	18.85 (0.89)	1.50 (1.26)	1.19	23	0.24
3	20.96 (1.08)	19.66 (1.12)	1.30 (1.56)	0.83	23	0.41
4	20.14 (0.83)	18.49 (0.77)	1.65 (1.13)	1.47	23	0.16
5	22.25 (1.38)	22.93 (1.13)	-0.69 (1.79)	-0.38	23	0.70
6	18.70 (0.90)	18.23 (0.78)	0.47 (1.19)	0.39	23	0.70
7	20.36 (0.92)	18.86 (0.93)	1.50 (1.31)	1.15	23	0.26

Chapter 2

Exact Type I Error in the Two Stage Analysis of Cluster Means

2.1 Introduction

The term "clustered data" commonly refers to data collected on individuals who are nested within a specific geographical or civil unit, e.g., children within schools, employees within worksites, or patients within physician practices. Clustered designs are often intentionally used to study the relationship of characteristics at the individual and cluster level on the response of interest. Many public health studies also require the use of a clustered instead of fully independent data collection designs due to logistical, ethical and cost constraints. For randomized studies, the trials through which clustered data arise are usually called group randomized trials, cluster randomized trials, or more generally, community trials [7, 24]. Such trials have been performed broadly across areas of medicine and public health, most notably in the areas of smoking prevention, physical activity promotion, occupational safety, nutrition, dentistry, and health policy. Specific features of clustered data are: the independent sampling unit is the cluster; characteristics of individuals within a cluster tend to be correlated, equally among each other; and the explanatory variable of primary scientific interest, e.g., treatment group, is applied at the cluster level, while data are collected at the individual or within-cluster level.

Many continuous outcomes of interest in public health studies with clustered data have an approximate Gaussian distribution. In this paper, the analysis of Gaussian clustered data within the framework of a linear model with Gaussian errors is discussed. For simplicity, in this paper, we refer to the explanatory variable of interest as treatment group, though discus-

sion about analyses for difference in treatment groups applies to any cluster level explanatory variable.

For data with one level of clustering, if data are balanced so that each cluster contributes the same number of observations, and if data have a common within-cluster correlation and individual error variance across treatment groups, then the set of outcome cluster means are sufficient statistics for inference about treatment group means. That is to say that knowledge of the individual level outcome data gives no additional information about the treatment means over that given by the outcome cluster means. This is true even when the outcome of interest depends on additional covariates other than treatment group, so long as the relationship between the outcome and covariates is the same across treatment groups. With the addition of covariates other than treatment group, knowledge of outcome cluster means and cluster covariate averages suffices for inference about treatment group. Sufficiency of cluster averages for inference about treatment groups is due to the special covariance structure of clustered data.

Because of this special feature, analysis of clustered data is often performed via a "two-stage" model, whereby the cluster means are computed first, often adjusted for covariates through a preliminary model excluding treatment group, and cluster means are the values of the response in a linear model at the second stage [24]. In a review of group randomized trials published in the American Journal of Public Health and Preventive Medicine from 1998 to 2002, Varnell *et al.* [35] showed that of the 47 trials that employed at least one statistical analysis appropriate for group randomized trials, 15 (32%) analyzed cluster means or another summary statistic.

For clustered data, $\mathcal{V}\left(\overline{y}_{hi}\right) = \left(\sigma_y^2/n_{hi}\right)\left[1 + \left(n_{hi} - 1\right)\rho\right]$; that is, the variance of each cluster mean is a function of the within cluster correlation, ρ, the individual error variance, σ_y^2, and the number of observations in the i-th cluster from the h-th treatment group, n_{hi}. Assuming homogeneous correlation and error variance across all clusters, cluster means have equivalent variances for balanced cluster sizes. This property of homogeneity of variances with the independence of clusters by the randomization scheme means that balanced Gaussian cluster means meet the assumptions of the familiar general linear univariate model with Gaussian errors. Because of this, for balanced data, a uniformly most powerful size-α test for the fixed effects exists and has exact null and non-null $\mathcal{F}$ distributions.

When data are unbalanced, cluster means no longer have homogenous variances, and thus

violate an assumption of the general linear univariate model. For unbalanced data, researchers often analyze a weighted univariate linear model to compensate for the heterogeneous variance of cluster means due to varying cluster sizes. Optimal weights for such a weighted linear model are equal to the inverse of the variances of the cluster means; however, such weights involve unknown variance parameters which must be estimated, so that at best, only approximately optimal weights are available. Further, no closed form expressions for maximum likelihood estimates of the variance parameters exist, so that estimates of variance parameters must be computed via iterative methods. Researchers often choose weights that avoid estimation of variance parameters, but are still approximately optimal.

Departures from optimal weights can inflate or deflate type I error for hypothesis tests for fixed effects in the weighted univariate linear model with Gaussian errors. In this paper, we focus on two weighting schemes: analysis of means weighted by cluster size and analysis of unweighted means. The main objective of this paper is to study the probability of type I error for a test of treatment difference with these two weighting schemes under the violation of the homogeneity of variance assumption in a general linear univariate model. To do this, we present a theorem which allows exact computation of type I error in the general linear univariate model with Gaussian errors under violation of covariance assumptions and then perform an enumeration study of type I error for several scenarios of cluster size imbalance.

2.2 Hypothesis Testing for Cluster Means

2.2.1 Notation

Discussion in this paper makes uses of matrix and random variable notation throughout. Lower case bold indicates a (column) vector, upper case bold, a matrix. Upper case italics indicates a non-matrix random variable. Matrix notation dominates over random variable notation, so that randomness of a matrix must be inferred from context. Muller and Stewart [23] gave a review of standard matrix operators used. In particular, this paper makes of use the direct sum operator, $\bigoplus_{i=1}^{I} \boldsymbol{A}_i = \mathrm{Dg}\left\{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_I\right\}$, which creates a block diagonal matrix, as well as the direct (or Kronecker) product: $\boldsymbol{A} \otimes \boldsymbol{B} = \{a_{ij}\boldsymbol{B}\}$ where a_{ij} is the element in the i-th row and j-th column of matrix $\boldsymbol{A}$.

Let $\boldsymbol{x} \sim \mathcal{N}_N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ indicate that the vector $\boldsymbol{x}$ $(N \times 1)$ follows an N-variate normal dis-

tribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $X \sim \mathcal{F}(\nu_1, \nu_2, \omega)$ indicate the random variable X had a noncentral F distribution with ν_1 numerator degrees of freedom, ν_2 denominator degrees of freedom, and noncentrality ω. Let $X \sim \chi^2(\nu_1, \omega)$ indicate that X has a non-central chi-square distribution with ν_1 degrees of freedom and noncentrality ω. With zero noncentralities, both the noncentral F and χ^2 distributions reduce to central versions. Kotz *et al.* [17] gave detailed information about these distributions.

This paper discusses hypothesis tests for difference in treatment group for data with one level of clustering. Table 2.1 describes all notation used in this paper to describe the structure of clustered data. In particular, we denote the number of treatment groups, clusters per treatment group, and observations per cluster with g, m_h, and n_{hi}, respectively.

2.2.2 Model and Hypothesis Statement

All theory discussed in this paper is in the context of a general linear univariate model with Gaussian errors. Muller and Stewart [23] or Muller and Fetterman [21], among others, gave detailed information about this type of model. Specify the model as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{2.1}$$

where $\boldsymbol{y}$ is the $(m \times 1)$ vector of cluster means, $\boldsymbol{X}$ is the $(m \times g)$ design matrix for the fixed effects with $\text{rank}(\boldsymbol{X}) = r$, and $\boldsymbol{e}$ is the $(m \times 1)$ vector of random errors. The vector $\boldsymbol{y}$ can also be written as $\boldsymbol{y} = \left\{\overline{y}_{11}, \ldots, \overline{y}_{1m_h}, \ldots, \overline{y}_{h1}, \ldots, \overline{y}_{hm_h}\right\}'$. We assume $\boldsymbol{e} \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{\Sigma})$, so that $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathcal{V}(\boldsymbol{y})$ is described below.

Because the level of the cluster is the independent sampling unit, observations in $\boldsymbol{y}$ are independent and $\boldsymbol{\Sigma}$ is diagonal. The exchangeable sampling scheme of clustered data naturally leads to an assumption of compound symmetric covariance structure for $\boldsymbol{\Sigma}$, with $\mathcal{V}(y_{hij}) = \sigma_y^2$ $\forall h, i, j$ and $\text{corr}\left(y_{hij}, y_{hij'}\right) = \rho$ for all observations $j \neq j'$. Such a structure for observations within a cluster implies that all observations have the same variance, σ_y^2, and pairs of observations with a cluster have the same correlation, ρ. Using these descriptions, the variance of each cluster mean can be shown to be:

$$\mathcal{V}(\overline{y}_{hi}) = \left(\sigma_y^2 / n_{hi}\right)\left[1 + (n_{hi} - 1)\,\rho\right]. \tag{2.2}$$

Let σ_{hi}^2 denote this expression for $\mathcal{V}(\overline{y}_{hi})$ so that $\boldsymbol{\Sigma} = \bigoplus_{h=1}^{g} \bigoplus_{i=1}^{m_h} \sigma_{hi}^2$.

We study the two sided general linear hypothesis (GLH) $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, where $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$, a linear combination of elements of $\boldsymbol{\beta}$. In most hypotheses of interest, $\boldsymbol{\theta}_0 = \mathbf{0}$. Such a hypothesis test describes differences in the fixed effects only. If $\boldsymbol{\theta}$ is estimable, requiring $\boldsymbol{C}$ $(a \times 1)$ to have full row rank [rank$(\boldsymbol{C}) = a$] ensures the GLH is testable. $\boldsymbol{\theta}$ is estimable if and only if $\boldsymbol{C} = \boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-}\left(\boldsymbol{X}'\boldsymbol{X}\right)$. Note that this requirement is automatically satisfied if $\boldsymbol{X}$ is full rank, so that $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ exists [21].

2.2.3 Hypothesis Testing for Cluster Means with Balanced Data

When data are balanced, $n_{hi} \equiv n$ for all h, i, so that $\sigma^2_{hi} = \sigma^2$ for all h, i. That is, for balanced data, all cluster means have the same variance, and $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_m$. Independence, normality, and homogeneity of errors of the cluster means meet the assumptions of the general linear univariate model.

In the general linear univariate model, the best linear unbiased estimator for the fixed effects, $\boldsymbol{\beta}$, is:

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

In turn, the best linear unbiased estimator for the contrasts $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\theta}} = \boldsymbol{C}\widehat{\boldsymbol{\beta}}$. Since $\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta},\ \sigma^2\boldsymbol{I}_m\right)$, it follows that $\widehat{\boldsymbol{\beta}} \sim \mathcal{N}\left[\boldsymbol{\beta},\ \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2\right]$ and $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}\left[\boldsymbol{\theta},\ \boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\sigma^2\right]$. Also, the reduced maximum likelihood estimator for σ^2 is:

$$\widehat{\sigma}^2 = \boldsymbol{y}'\left[\boldsymbol{I}_m - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}\right]\boldsymbol{y}/\left(m - r\right).$$

Using quadratic form theory for normal vectors, $\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)'\left[\mathcal{V}\left(\widehat{\boldsymbol{\theta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)/\sigma^2$ $\sim \chi^2\left(a,\, \omega\right)$ independently of $\left(m - r\right)\widehat{\sigma}^2/\sigma^2 \sim \chi^2\left(m - r\right)$, so that:

$$T_u = \left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)'\left[\mathcal{V}\left(\widehat{\boldsymbol{\theta}}\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)/a\right]/\widehat{\sigma}^2 \sim \mathcal{F}\left(a, m - r, \omega\right).$$

The noncentrality factor ω is given by $\omega = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)/\sigma^2$. T_u can be shown to provide a uniformly most powerful size α test for the general linear hypothesis.

Algebra expresses T_u as:

$$T_u = \frac{\left(\boldsymbol{y} - \boldsymbol{a}_c\right)'\boldsymbol{A}_h\left(\boldsymbol{y} - \boldsymbol{a}_c\right)/a}{\left(\boldsymbol{y} - \boldsymbol{a}_c\right)'\boldsymbol{A}_e\left(\boldsymbol{y} - \boldsymbol{a}_c\right)/\left(m - r\right)} \tag{2.3}$$

where $\boldsymbol{a}_c = \boldsymbol{X}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1}\boldsymbol{\theta}_0$, $\boldsymbol{A}_h = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$, and $\boldsymbol{A}_e = \boldsymbol{I}_m - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}$. This form of T_u is useful, because it expresses both the numerator

and denominator as quadratic forms in the same vector, necessary for application of theory developed later in this paper.

2.2.4 Extension to Weighted Model

The general linear univariate model assumes $\boldsymbol{\Sigma}$ is proportional to an identity matrix, that is $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$. An optimal test for the general linear hypothesis can still be derived with the less restrictive assumption that $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{W}^{-1}$, that is, that the covariance matrix is known up to a constant weight matrix $\boldsymbol{W}$ [19]. In this case, a uniformly most powerful size-α test for the fixed effects is given by:

$$T_w = \frac{(\boldsymbol{y} - \boldsymbol{a}_c)' \boldsymbol{A}_{hw} (\boldsymbol{y} - \boldsymbol{a}_c) / a}{(\boldsymbol{y} - \boldsymbol{a}_c)' \boldsymbol{A}_{ew} (\boldsymbol{y} - \boldsymbol{a}_c) / (m - r)}$$

where $\boldsymbol{A}_{hw} = \boldsymbol{W}\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} \boldsymbol{C}' \left[\boldsymbol{C} (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} \boldsymbol{C}'\right]^{-1} \boldsymbol{C} (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{W}$ and $\boldsymbol{A}_{ew} = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{W}$. Following similar arguments as in the previous section, $T_w \sim \mathcal{F}(a,\, m - r,\, \omega_w)$, where the noncentrality in the weighted model is:

$$\omega_w = (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{\theta}_0)' \left[\boldsymbol{C} \left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1} \boldsymbol{C}'\right]^{-1} (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{\theta}_0) / \sigma^2.$$

2.2.5 Hypothesis Testing for Cluster Means with Unbalanced Data

The exact distribution of the test statistic T_w depends on the assumption that $\boldsymbol{\Sigma}$ can be written in the form $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{W}^{-1}$. When this doesn't hold this statistic has only an approximate $\mathcal{F}$ distribution.

For unbalanced clustered data where $\boldsymbol{\Sigma} = \bigoplus_{h=1}^{g} \bigoplus_{i=1}^{m_h} \sigma_{hi}^2$, $\boldsymbol{\Sigma}$ does not have this form, since the variance components are unknown. The common strategy is to choose a $\boldsymbol{W}$ such that $\boldsymbol{\Sigma} \approx \sigma^2 \boldsymbol{W}^{-1}$.

If ρ is small and cluster sizes are small, $\boldsymbol{\Sigma} \approx \sigma_y^2 \bigoplus_{h=1}^{a} \bigoplus_{i=1}^{m_h} (1/n_{hi})$, so that a sensible choice is $\boldsymbol{W} = \bigoplus_{h=1}^{a} \bigoplus_{i=1}^{m_h} (n_{hi})$. If cluster sizes are not highly variable, $\boldsymbol{\Sigma} \approx \sigma^2 \boldsymbol{W}^{-1}$ directly, so that data analysts may choose an unweighted means approach with $\boldsymbol{W} = \boldsymbol{I}_m$. Research is needed to investigate the impact of these approximate weights on the type I error rate of a hypothesis test for fixed effects in the weighted general linear univariate model, when applied to unbalanced clustered data.

2.3 Theoretical Result to Compute Probabilities Under Violation of Assumptions

To study type I error in the general linear univariate model under violation of the homogeneity of variance assumption one can perform a series of simulations, which can take considerable computation time, or employ approximations in the style of Satterthwaite [28], which naturally have inherent imprecision.

Using theory outlined in several sources, e.g., Box [1] and Johnson and Kotz [11, Ch. 29], Theorem 1 below gives a method to exactly enumerate rather than simulate test size for the hypothesis test for fixed effects in the general linear univariate model. A single computation of type I error takes seconds, several thousand take minutes, as opposed to hours or days required for a simulation study.

Theorem 1

Suppose $\boldsymbol{y} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is any known covariance matrix. Let $\boldsymbol{\Phi}$ be the Cholesky factor of $\boldsymbol{\Sigma}$ so that $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Phi}'$. Define the test statistic:

$$T = \frac{(\boldsymbol{y} - \boldsymbol{c})' \boldsymbol{A}_1 (\boldsymbol{y} - \boldsymbol{c}) / \nu_1}{(\boldsymbol{y} - \boldsymbol{c})' \boldsymbol{A}_2 (\boldsymbol{y} - \boldsymbol{c}) / \nu_2},$$

where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are any known, constant matrices. Also define $\boldsymbol{A}_3 = \boldsymbol{\Phi}' (\boldsymbol{A}_1/\nu_1 - f\boldsymbol{A}_2/\nu_2) \boldsymbol{\Phi}$, where $f = \mathcal{F}^{-1}(1 - \alpha, \nu_1, \nu_2)$. Express $\boldsymbol{A}_3$ as its spectral decomposition $\boldsymbol{A}_3 = \boldsymbol{V}\mathrm{Dg}(\boldsymbol{\lambda})\boldsymbol{V}'$, where $\boldsymbol{\lambda}$ and $\boldsymbol{V}$ are the vector of eigenvalues and matrix of eigenvectors of $\boldsymbol{A}_3$, respectively. Finally, define a random vector $\boldsymbol{x}$ such that such that $\boldsymbol{x}^2 \sim \chi^2(\boldsymbol{1}, \boldsymbol{\omega})$ with $\boldsymbol{\omega} = \left[\boldsymbol{V}'\boldsymbol{\Phi}^{-1}(\boldsymbol{\mu} - \boldsymbol{c})\right]^2$. Here, the square notation denotes squaring each element of the vector.

Then:

$$\mathrm{Prob}\{T \leq f\} = \mathrm{Prob}\left\{\sum_{i=1}^{N} \lambda_i x_i^2 \leq 0\right\},$$

where the $\{\lambda_i\}$ and $\{x_i\}$ are elements of $\boldsymbol{\lambda}$ and $\boldsymbol{x}$, respectively. That is, the CDF of T can be written as a sum of weighted central on non-central chi-square random variables with weights given in $\boldsymbol{\lambda}$ and noncentralities given in $\boldsymbol{\omega}$. With these weights and noncentralities, probabilities from the CDF of T can be computed using the algorithm in Davies [4]. Proof of Theorem 1 is given in the Appendix.

2.4 Description of Enumerations

The remainder of this paper describes an enumeration study to compute type I error for a hypothesis test of treatment difference in the analysis of cluster means in a weighted univariate linear model with Gaussian errors under the violation of the homogeneity of variance assumption due to varying cluster sizes.

2.4.1 Application of Theorem 1 to Clustered Data

Theorem 1 may be applied to calculate type I error for clustered data with the following steps:

1. Specify the model matrices $\boldsymbol{C}$, $\boldsymbol{X}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\theta}_0$ and target type I error, α.

2. Specify the matrix of weights, $\boldsymbol{W}$.

3. Compute constants $\nu_2 = m - \text{rank}\left(\boldsymbol{X}\right)$, $\nu_1 = \text{rank}\left(\boldsymbol{C}\right)$, and $f = \mathcal{F}^{-1}\left(1-\alpha, \nu_1, \nu_2\right)$.

4. Compute the matrices:

$$\boldsymbol{A}_1 = \boldsymbol{W}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}$$

$$\boldsymbol{A}_2 = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}$$

$$\boldsymbol{c} = \boldsymbol{X}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1}\boldsymbol{\theta}_0.$$

   These are equivalent to matrices $\boldsymbol{A}_{hw}$ and $\boldsymbol{A}_{ew}$ defined in section 2.2.4, and $\boldsymbol{a}_c$ in section 2.2.3, respectively.

5. Compute $\boldsymbol{\Phi}$, the Cholesky factor of $\boldsymbol{\Sigma}$, such that $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Phi}'$.

6. Compute the matrix $\boldsymbol{A}_3 = \boldsymbol{\Phi}'\left(\boldsymbol{A}_1/\nu_1 - f\boldsymbol{A}_2/\nu_2\right)\boldsymbol{\Phi}$.

7. Compute $\boldsymbol{\lambda}$ and $\boldsymbol{V}$, the vector of eigenvalues and matrix of eigenvectors of $\boldsymbol{A}_3$, respectively.

8. Use $\boldsymbol{\lambda}$ as the vector of weights in a module that perform Davies' algorithm, available at http://www.bios.unc.edu/∼muller. For type I error calculations, $\boldsymbol{\omega} = \boldsymbol{0}$.

A module that performs these calculations, CLUSMOD, may be downloaded for free off the above web site, including brief documentation for its use.

2.4.2 Description of Parameters of Imbalance

We study a two group comparison, since most analyses of clustered data involve two treatment arms [35]. We characterize imbalance in the number of observations per cluster with four parameters: the average cluster sizes of each treatment group, denoted as $\overline{n}_1$ and $\overline{n}_2$, and the ratio of maximum to minimum cluster sizes for each treatment group, denoted as r_1 and r_2. Though the theoretical development in this paper primarily focuses on imbalance in the number of observations per cluster across treatment groups, imbalance in the number of clusters affects type I error, sometimes dramatically when combined with imbalance in number of observations per cluster. Thus, the number of clusters per treatment group, denoted as m_1 and m_2, were also varied. The fifth and final parameter varied was the within cluster correlation, ρ. All computations assumed target $\alpha = 0.05$.

2.4.3 Values of Parameters of Imbalance

Values for each of ρ, m_1, m_2, $\overline{n}_1$ and $\overline{n}_2$ were chosen as follows to best represent scenarios of non-randomized clustered designs in the literature, which often differ from randomized studies in that they more often have unbalanced number of clusters per treatment group. This enumeration studied $m_1, m_2 \in \{2, 4, 8, 16, 32\}$, $\overline{n}_1, \overline{n}_2 \in \{8, 16, 32, 64, 128, 256\}$, $r_1, r_2 \in \{1, 2, 4, 8\}$ and $\rho \in \{0.001, 0.01, 0.1\}$. Any design with $r_1 = r_2 = 1$ and $\overline{n}_1 = \overline{n}_2$ has balanced cluster sizes for all clusters and so has type I error equal to 0.05 exactly; all others are unbalanced designs. Davies [4] algorithm did not converge for many of the $m_1 = 2$, $m_2 = 2$ cases, so this combination of number of clusters per treatment group was omitted.

A full factorial combination of these parameters yields 41,472 cases of imbalance. Many of these lead to the same design with respect to type I error computation. For example, a case with $m_1 = m_2 = 4$, $\overline{n}_1 = \overline{n}_2 = 8$, $r_1 = 1$ and $r_2 = 2$ will have the same type I error as a case with r_1 and r_2 reversed. Unique cases can be characterized as those which have $(m_1 < m_2)$, or $(m_1 = m_2$ and $\overline{n}_1 < \overline{n}_2)$, or $(m_1 = m_2$, $\overline{n}_1 = \overline{n}_2$, and $r_1 \leq r_2)$. Of the 41,472 cases, 20,880 are unique. Type I error was computed only for unique cases.

2.4.4 Generation of Cluster Sizes

Cluster sizes were selected from a Gaussian distribution with mean $\overline{n}_h$, the specified average cluster size for treatment group h; the standard deviation of the Gaussian distribution was computed as follows. Define $n_{h,max}$ and $n_{h,min}$ as the maximum and minimum cluster sizes, respectively, in treatment group h. Since the Gaussian distribution is symmetric, $\overline{n}_h = (n_{h,max} + n_{h,min})/2$. By definition of the enumeration study parameters, $n_{h,max} = r_h n_{h,min}$. Substituting this back into the previous expression and solving for $n_{h,max}$ and $n_{h,min}$ yields $n_{h,min} = 2_h/(r_h + 1)$ and $n_{h,max} = 2_h r_h/(r_h + 1)$. Since $\approx 95\%$ percent of the Gaussian distribution in within two standard deviations of the mean, we fixed $n_{h,min}$ and $n_{h,max}$ at two standard deviations from mean, so that the standard deviation, σ, can then be calculated as $\sigma = (n_{h,max} - n_{h,min})/4 = 2_h(r_h - 1)/4(r_h + 1)$. Thus, cluster sizes were given the distribution:

$$n_{hi} \sim \mathcal{N}\left\{\overline{n}_h,\ [2_h r_h/(r_h + 1)]^2\right\}.$$

Define $c_2 = \text{Prob}\{(Z < -2)\}$, where $Z \sim \mathcal{N}(0, 1)$. Cluster sizes $\{n_{hi}\}$ were then computed such that:

$$n_{hi} = \mathcal{Z}^{-1}[c_2 + (i - 1)(1 - c_2)/m_h]\,\sigma + \overline{n}_h$$

for $h = 1, 2$ and $i = 1, \ldots, m_h$, where $\mathcal{Z}^{-1}(p)$ denotes a function with returns the p-th quantile from a standard normal distribution. Cluster sizes were not rounded so that actual ratios for maximum to minimum cluster size were achieved.

2.5 Results of Enumeration Study

2.5.1 Overview

The purpose of this enumeration study is to evaluate which values of each of number of clusters per treatment group, number of observations per cluster, ratio of maximum to minimum cluster size, and within-cluster correlation affect Type I error from nominal levels. For the discussion that follows, we define type I error, α, to be approximately unbiased if it varies from 0.05 by less than a multiplicative factor of 2; that is, if $0.025 \leq \alpha \leq 0.1$.

Discussion of these enumeration results is complicated as this study enumerated Type I error for over 20,000 cases of imbalance. Further, type I error cannot be derived as an explicit function

of the parameters of imbalance, so summarization over imbalance factors or combinations of factors yields at best only general conclusions.

2.5.2 Main Effects of Imbalance Parameters

Tables 2.2 - 2.4 give descriptive statistics for type I error over all cases and for the main effects of each parameter of imbalance and ratios of them, where applicable. No one parameter of imbalance by itself was a major predictor of type I error, so only few conclusions may be drawn from displays of type I error for the main effect of each imbalance parameter.

2.5.2.1 Within Cluster Correlation

Section 2 of Table 2.2 displays type I error by ρ, where $\rho \in \{.001, .01, .1\}$. Regardless of the imbalance in number of clusters or cluster sizes, the analysis of means weighted by cluster size was approximately unbiased when $\rho = .001$. Though sometimes still biased, the analysis of unweighted means was approximately unbiased for many more cases than the analysis with cluster size weights when $\rho = .1$.

As ρ increased, type I error for the analysis of means weighted by cluster size became more biased. For the analysis of unweighted means, as ρ increased, type I error became less biased, so that the two weighting schemes show opposite relationships between type I error and ρ.

2.5.2.2 Number of Clusters Per Treatment Group

Sections 3-6 of Table 2.2 display type I error for combinations of $m_1 \times m_2$ and their ratio, m_2/m_1, where $m_1, m_2 \in \{2, 4, 8, 16, 32\}$. Descriptive statistics are similar for combinations of $m_1 \times m_2$ with the same ratio, so combinations of $m_1 \times m_2$ are displayed in order of m_2/m_1.

When $m_1 = m_2$, that is, when the number of clusters per treatment group is equal, the analysis of unweighted means provided approximately unbiased type I error regardless of ρ or imbalance in cluster sizes. No other combinations or ratios of number of clusters uniformly lead to approximately unbiased type I error in the analysis of unweighted means, nor did any combinations or ratios of number of clusters do so for the analysis of means weighted by cluster size.

In general, as m_2/m_1 increased, type I error became more biased for both weightings. As

the number of clusters increased for a given ratio of m_2/m_1, type I error became less biased.

2.5.2.3 Average Cluster Size Per Treatment Group

Table 2.3 displays type I error for combinations of $\overline{n}_1 \times \overline{n}_2$ and their ratio, $\overline{n}_2/\overline{n}_1$, where $\overline{n}_1, \overline{n}_2 \in \{8, 16, 32, 64, 128, 256\}$. As with number of clusters in the previous section, descriptive statistics are similar for combinations of $\overline{n}_1 \times \overline{n}_2$ with the same ratio, so combinations of $\overline{n}_1 \times \overline{n}_2$ are displayed in order of $\overline{n}_2/\overline{n}_1$.

Both weightings controlled type I error well for values of almost all other factor of imbalance when $\overline{n}_1 = \overline{n}_2$, that is, when the average cluster size of the two treatment groups was equal.

For all combinations of $\overline{n}_1$ and $\overline{n}_2$, as average cluster size increased, type I error because less biased for the analysis of unweighted means and more biased for the analysis of means weighted by cluster size. As $\overline{n}_2/\overline{n}_1$ increased, the analysis of unweighted means more often lead to anti-conservative type I error and the analysis of means weighted by cluster size more often lead to conservative type I error.

2.5.2.4 Ratio of Maximum to Minimum Cluster Size

Table 2.4 displays type I error for combinations of $r_1 \times r_2$ and their ratio, r_2/r_1, where r_1, $r_2 \in \{1, 2, 4, 8\}$. The ratio of maximum to minimum cluster size was the least important factor of imbalance in determining type I error, as combinations of $r_1 \times r_2$ had largely similar type I error and did not vary significantly with respect to type I error when stratified by any of the other parameters of imbalance (not shown). As such, further discussions and displays will summarize results over r_1 and r_2.

2.5.3 Combinations of Number of Clusters, Average Cluster Size, and Within Cluster Correlation

Variation in type I error as a function of parameters of imbalance is best explained in groups of $\rho \times m_1 \times m_2 \times \overline{n}_1 \times \overline{n}_2$; however, 1,332 such groups exist, so that descriptive statistics for every group cannot be displayed compactly.

Tables 2.5 and 2.6 give type I error for combinations of $m_2/m_1 \times \overline{n}_2/\overline{n}_1$, the ratios of number of clusters and average cluster size, respectively. As mentioned previously, when $m_2/m_1 = 1$

and $\overline{n}_2/\overline{n}_1 = 1$, that is, when both the number of clusters and average cluster size per treatment group are equal, hypothesis tests for both weightings are unbiased. In this case, type I error for the analysis of unweighted means is always conservative and is always anti-conservative for the analysis of means weighted by cluster size.

When $m_2 > m_1$ and $\overline{n}_2 > \overline{n}_1$, that is, when one treatment group has more clusters and more observations per cluster than the other, the analysis of unweighted means tends to be anti-conservative and the analysis of means weighted by cluster size tends to be conservative. The opposite is true when $m_2 > m_1$ and $\overline{n}_2 < \overline{n}_1$, that is when one treatment group has more clusters and the other has more observations per cluster. In this case, the analysis of unweighted means tends to be conservative and the analysis of means weighted by cluster size tends to be anti-conservative.

Figures 2.1 and 2.2 show for which combinations of $\rho \times m_1 \times m_2 \times \overline{n}_1 \times \overline{n}_2$ each weighted analysis is unbiased. Finally, Figure 2.3 shows for which combinations of $\rho \times m_1 \times m_2 \times \overline{n}_1 \times \overline{n}_2$ both weighting schemes lead to unbiased type I error, when each of unweighted means or weighting means by cluster size is preferred, and when both are biased.

2.6 Conclusions and Recommendations

Theory developed in this paper gives a way to enumerate Type I error exactly. This is a useful diagnostic tool for examining type I and type II error for hypothesis tests in the general linear model with Gaussian errors when assumptions of independence and homogeneity have been violated. The enumeration study presented here showed that with equal number of clusters per treatment group an analysis of unweighted means is recommended for values of correlation and cluster size imbalance found in most clustered data settings. In turn, when within cluster correlation is small, an analysis of means weighted by cluster size is recommended. Type I error is a complicated function of imbalance parameters and is robust to moderate imbalance in either the number of clusters or number of observations per cluster. Researchers should avoid imbalance in both of these quantities in order to avoid significant bias in type I error for hypothesis tests in the analysis of weighted cluster means.

2.7 Proof

Proof of Theorem 1:

First express Prob$\{T \leq f\}$ as a sum of quadratic forms:

$$\begin{aligned}\text{Prob}\left\{T \leq f\right\} &= \text{Prob}\left\{\frac{\left(\boldsymbol{y}-\boldsymbol{c}\right)'\boldsymbol{A}_1\left(\boldsymbol{y}-\boldsymbol{c}\right)/\nu_1}{\left(\boldsymbol{y}-\boldsymbol{c}\right)'\boldsymbol{A}_2\left(\boldsymbol{y}-\boldsymbol{c}\right)/\nu_2} \leq f\right\} \\ &= \text{Prob}\left\{\left(\boldsymbol{y}-\boldsymbol{c}\right)'\left(\boldsymbol{A}_1/\nu_1 - f\boldsymbol{A}_2/\nu_2\right)\left(\boldsymbol{y}-\boldsymbol{c}\right) \leq 0\right\}\end{aligned}$$

This expresses Prob$\{T \leq f\}$ as a sum of weighted squared Gaussian random variables (the elements of $\boldsymbol{y}-\boldsymbol{c}$), but they are not independent. Now express $\boldsymbol{y}$ in terms of a transformation of $\boldsymbol{z}$, a vector of independent random variables, as $\boldsymbol{y} = \boldsymbol{\Phi}\left(\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\boldsymbol{\mu}\right)$ and substitute this back into the previous expression for the CDF of T:

$$\begin{aligned}\text{Prob}\left\{T \leq f\right\} &= \text{Prob}\left\{\left[\boldsymbol{\Phi}\left(\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\boldsymbol{\mu}\right) - \boldsymbol{c}\right]'\left(\boldsymbol{A}_1/\nu_1 - f\boldsymbol{A}_2/\nu_2\right)\left[\boldsymbol{\Phi}\left(\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\boldsymbol{\mu}\right) - \boldsymbol{c}\right] \leq 0\right\} \\ &= \text{Prob}\left\{\left[\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\boldsymbol{\mu} - \boldsymbol{\Phi}^{-1}\boldsymbol{c}\right]'\boldsymbol{A}_3\left[\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\boldsymbol{\mu} - \boldsymbol{\Phi}^{-1}\boldsymbol{c}\right] \leq 0\right\} \\ &= \text{Prob}\left\{\left[\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right]'\boldsymbol{A}_3\left[\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right] \leq 0\right\}\end{aligned}$$

Now we have expressed Prob$\{T \leq f\}$ as a sum of weighted squared independent Gaussian, i.e., chi-square, random variables, but we need to know the weights. We can show that $\boldsymbol{A}_3$ is symmetric, so that we can write its' spectral decomposition into a matrix of eigenvectors, $\boldsymbol{V}$, and vector of eigenvalues, $\boldsymbol{\lambda}$, as $\boldsymbol{A}_3 = \boldsymbol{V}\text{Dg}(\boldsymbol{\lambda})\boldsymbol{V}'$. By definition, $\boldsymbol{V}$ is orthonormal so that $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_m$. Substitute this back into the previous expression for $\boldsymbol{A}_3$ to give:

$$\begin{aligned}\text{Prob}\left\{T \leq f\right\} &= \text{Prob}\left\{\left[\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right]'\boldsymbol{V}\text{Dg}\left(\boldsymbol{\lambda}\right)\boldsymbol{V}'\left[\boldsymbol{z} + \boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right] \leq 0\right\} \\ &= \text{Prob}\left\{\left[\boldsymbol{V}'\boldsymbol{z} + \boldsymbol{V}'\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right]'\text{Dg}\left(\boldsymbol{\lambda}\right)\left[\boldsymbol{V}'\boldsymbol{z} + \boldsymbol{V}'\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right] \leq 0\right\} \\ &= \text{Prob}\left\{\boldsymbol{x}'\text{Dg}\left(\boldsymbol{\lambda}\right)\boldsymbol{x} \leq 0\right\} \\ &= \text{Prob}\left\{\sum_{i=1}^{m}\lambda_i x_i^2 \leq 0\right\},\end{aligned}$$

where $\boldsymbol{x} = \boldsymbol{V}'\boldsymbol{z} + \boldsymbol{V}'\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)$ and $\boldsymbol{x} \sim \mathcal{N}_m\left[\boldsymbol{V}'\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right), \boldsymbol{I}_m\right]$, so that $\boldsymbol{x}^2 \sim \chi^2\left(1, \boldsymbol{\omega}\right)$ with $\boldsymbol{\omega} = \left[\boldsymbol{V}'\boldsymbol{\Phi}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{c}\right)\right]^2$. □

Table 2.1: Summary of Non Matrix Notation

Symbol	Definition
	Indicies
$h = 1, \ldots, g$	Indexes treatment groups
$i = 1, \ldots, m_h$	Indexes clusters within treatment group
$j = 1, \ldots, n_{hi}$	Indexes observations within cluster
	Numbers of Clusters and Observations
g	Number of treatment groups
m_h	Number of clusters in treatment group h
$m = \sum_{h=1}^{g} m_h$	Total number of clusters
n_{hi}	Number of observations within a cluster when cluster sizes are unequal
n	Number of observations within a cluster when cluster sizes are equal
$n_h = \sum_{i=1}^{m_h} n_{hi}$	Number of observations in treatment group h
$N = \sum_{h=1}^{g} \sum_{i=1}^{m_h} n_{hi}$	Total number of observations
	Outcome Notation
y_{hij}	Outcome for observation j of cluster i in treatment group h
$\overline{y}_{hi} = \frac{1}{n_{hi}} \sum_{j=1}^{n_{hi}} y_{hij}$	Outcome mean for cluster i in treatment group h
$\overline{y}_h = \frac{1}{m_h} \sum_{i=1}^{m_h} \sum_{j=1}^{n_{hi}} y_{hij}$	Outcome mean for treatment group h

Table 2.2: Type I error over all cases and by ρ, $m_1 \times m_2$ and m_2/m_1

		Means Unweighted								Means Weighted by Cluster Size							
		N	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]	
All		20,880	<.001	.034	.050	.080	.725	.725	12,830 (61%)	<.001	.035	.050	.057	.484	.484	15,389	(74%)
$\rho =$	.001	6,960	<.001	.022	.051	.115	.725	.725	3,062 (44%)	.030	.046	.050	.051	.079	.049	6,960	(100%)
	.01	6,960	<.001	.029	.051	.096	.614	.614	3,757 (54%)	<.001	.029	.051	.059	.239	.239	5,133	(74%)
	.1	6,960	.006	.043	.050	.061	.267	.261	6,011 (86%)	<.001	.012	.049	.078	.484	.484	3,296	(47%)
$m_1, m_2 =$	4,4	900	.038	.049	.050	.054	.082	.044	900 (100%)	.004	.040	.049	.051	.091	.087	790	(88%)
	8,8	900	.046	.050	.050	.053	.066	.020	900 (100%)	.001	.032	.047	.050	.074	.073	734	(82%)
	16,16	900	.049	.050	.050	.052	.058	.009	900 (100%)	<.001	.028	.046	.050	.068	.068	712	(79%)
	32,32	900	.050	.050	.050	.051	.054	.004	900 (100%)	<.001	.026	.046	.050	.065	.065	692	(77%)
$m_1, m_2 =$	2,4	1,728	.017	.035	.050	.070	.250	.233	1,328 (77%)	.002	.044	.051	.059	.153	.151	1,489	(86%)
	4,8	1,728	.012	.034	.050	.075	.201	.188	1,208 (70%)	<.001	.037	.049	.053	.106	.105	1,455	(84%)
	8,16	1,728	.010	.032	.050	.076	.179	.169	1,175 (68%)	<.001	.034	.048	.052	.090	.090	1,398	(81%)
	16,32	1,728	.008	.031	.050	.076	.168	.160	1,175 (68%)	<.001	.031	.047	.052	.084	.084	1,382	(80%)
$m_1, m_2 =$	2,8	1,728	.002	.026	.052	.107	.459	.458	834 (48%)	<.001	.038	.051	.070	.251	.251	1,197	(69%)
	4,16	1,728	.001	.023	.051	.104	.380	.379	811 (47%)	<.001	.035	.051	.063	.174	.174	1,239	(72%)
	8,32	1,728	<.001	.021	.051	.103	.339	.338	800 (46%)	<.001	.033	.050	.059	.150	.150	1,252	(72%)
$m_1, m_2 =$	2,16	1,728	<.001	.020	.056	.141	.641	.641	654 (38%)	<.001	.034	.052	.083	.374	.374	1,042	(60%)
	4,32	1,728	<.001	.017	.052	.127	.534	.534	658 (38%)	<.001	.032	.051	.075	.282	.282	1,047	(61%)
$m_1, m_2 =$	2,32	1,728	<.001	.017	.059	.165	.725	.725	587 (34%)	<.001	.033	.052	.091	.484	.484	960	(56%)
$m_2/m_1 =$	1	3,600	.038	.050	.050	.052	.082	.044	3,600 (100%)	<.001	.033	.047	.050	.091	.091	2,928	(81%)
	2	6,912	.008	.033	.050	.074	.250	.241	4,886 (71%)	<.001	.036	.049	.053	.153	.153	5,724	(83%)
	4	5,184	<.001	.023	.051	.104	.459	.459	2,445 (47%)	<.001	.035	.051	.064	.251	.251	3,688	(71%)
	8	3,456	<.001	.019	.054	.135	.641	.641	1,312 (38%)	<.001	.033	.052	.079	.374	.374	2,089	(60%)
	16	1,728	<.001	.017	.059	.165	.725	.725	587 (34%)	<.001	.033	.052	.091	.484	.484	960	(56%)

[1] Range = Max - Min

[2] Number (%) of cases with $\alpha \in (.025, .10)$

Table 2.3: Type I error by $\bar{n}_1 \times \bar{n}_2$ and $\bar{n}_2/\bar{n}_1$

		Means Unweighted								Means Weighted by Cluster Size						
		N	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]
$\bar{n}_1, \bar{n}_2 =$	8,8	600	.035	.048	.050	.056	.183	.148	572 (95%)	.050	.050	.051	.053	.082	.032	600 (100%)
	16,16	600	.035	.048	.050	.055	.182	.147	579 (97%)	.050	.050	.051	.055	.094	.044	600 (100%)
	32,32	600	.035	.048	.050	.055	.180	.145	581 (97%)	.050	.050	.052	.057	.103	.053	594 (99%)
	64,64	600	.035	.049	.050	.054	.177	.142	583 (97%)	.050	.051	.053	.059	.110	.060	589 (98%)
	128,128	600	.036	.049	.050	.053	.171	.135	588 (98%)	.050	.051	.054	.062	.115	.065	587 (98%)
	256,256	600	.037	.050	.050	.052	.160	.123	592 (99%)	.050	.052	.056	.063	.119	.069	586 (98%)
$\bar{n}_1, \bar{n}_2 =$	8,16	1,152	.006	.033	.050	.080	.330	.325	827 (72%)	.020	.047	.050	.053	.153	.133	1,099 (95%)
	16,32	1,152	.006	.035	.050	.075	.328	.322	867 (75%)	.015	.046	.050	.056	.187	.172	1,042 (90%)
	32,64	1,152	.006	.039	.050	.070	.322	.316	903 (78%)	.012	.044	.050	.060	.213	.201	1,009 (88%)
	64,128	1,152	.007	.042	.050	.065	.312	.305	954 (83%)	.010	.043	.050	.064	.229	.219	967 (84%)
	128,256	1,152	.008	.044	.050	.060	.294	.286	1,001 (87%)	.009	.041	.050	.069	.239	.230	893 (78%)
$\bar{n}_1, \bar{n}_2 =$	8,32	1,152	<.001	.021	.051	.110	.472	.472	431 (37%)	.004	.037	.049	.053	.250	.246	924 (80%)
	16,64	1,152	.001	.025	.051	.100	.467	.466	572 (50%)	.002	.031	.048	.055	.301	.300	853 (74%)
	32,128	1,152	.001	.028	.051	.088	.456	.456	652 (57%)	.001	.025	.046	.059	.337	.336	717 (62%)
	64,256	1,152	.001	.032	.050	.077	.437	.436	696 (60%)	.001	.021	.043	.065	.359	.358	651 (57%)
$\bar{n}_1, \bar{n}_2 =$	8,64	1,152	<.001	.016	.053	.136	.587	.587	340 (30%)	<.001	.023	.046	.052	.350	.350	760 (66%)
	16,128	1,152	<.001	.021	.052	.120	.578	.578	459 (40%)	<.001	.016	.043	.053	.407	.407	650 (56%)
	32,256	1,152	<.001	.025	.052	.102	.562	.562	559 (49%)	<.001	.009	.038	.056	.445	.445	571 (50%)
$\bar{n}_1, \bar{n}_2 =$	8,128	1,152	<.001	.013	.053	.155	.670	.670	324 (28%)	<.001	.011	.040	.051	.428	.428	605 (53%)
	16,256	1,152	<.001	.019	.053	.132	.658	.658	429 (37%)	<.001	.005	.034	.052	.484	.484	549 (48%)
$\bar{n}_1, \bar{n}_2 =$	8,256	1,152	<.001	.012	.054	.165	.725	.725	321 (28%)	<.001	.003	.032	.046	.474	.474	543 (47%)
$\bar{n}_2/\bar{n}_1 =$	1	3,600	.035	.049	.050	.054	.183	.148	3,495 (97%)	.050	.050	.052	.058	.119	.069	3,556 (99%)
	2	5,760	.006	.038	.050	.071	.330	.325	4,552 (79%)	.009	.045	.050	.059	.239	.230	5,010 (87%)
	4	4,608	<.001	.026	.051	.099	.472	.472	2,351 (51%)	.001	.030	.048	.058	.359	.358	3,145 (68%)
	8	3,456	<.001	.020	.052	.125	.587	.587	1,358 (39%)	<.001	.017	.043	.054	.445	.445	1,981 (57%)
	16	2,304	<.001	.016	.053	.149	.670	.670	753 (33%)	<.001	.008	.036	.051	.484	.484	1,154 (50%)
	32	1,152	<.001	.012	.054	.165	.725	.725	321 (28%)	<.001	.003	.032	.046	.474	.474	543 (47%)

[1] Range = Max - Min

[2] Number (%) of cases with $\alpha \in (.025, .10)$

Table 2.4: Type I error by $r_1 \times r_2$ and r_2/r_1

		Means Unweighted							Means Weighted by Cluster Size						
	N	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]
$r_1, r_2 =$ 1,1	1,332	<.001	.034	.050	.074	.601	.601	835 (63%)	<.001	.032	.049	.051	.351	.351	990 (74%)
2,2	1,332	<.001	.034	.050	.075	.616	.616	827 (62%)	<.001	.033	.049	.054	.378	.378	992 (74%)
4,4	1,332	<.001	.033	.051	.078	.656	.656	819 (61%)	<.001	.036	.050	.060	.435	.435	994 (75%)
8,8	1,332	<.001	.036	.051	.086	.712	.712	807 (61%)	<.001	.039	.051	.065	.484	.484	974 (73%)
$r_1, r_2 =$ 1,2	2,592	<.001	.033	.050	.076	.618	.618	1,592 (61%)	<.001	.032	.049	.052	.378	.378	1,910 (74%)
2,4	2,592	<.001	.033	.050	.079	.662	.662	1,593 (61%)	<.001	.034	.049	.057	.434	.434	1,910 (74%)
4,8	2,592	<.001	.034	.051	.085	.719	.719	1,558 (60%)	<.001	.037	.050	.063	.484	.484	1,900 (73%)
$r_1, r_2 =$ 1,4	2,592	<.001	.033	.050	.079	.664	.664	1,613 (62%)	<.001	.034	.049	.055	.434	.434	1,911 (74%)
2,8	2,592	<.001	.035	.050	.085	.724	.724	1,583 (61%)	<.001	.036	.050	.059	.483	.483	1,903 (73%)
$r_1, r_2 =$ 1,8	2,592	<.001	.035	.050	.084	.725	.725	1,603 (62%)	<.001	.035	.050	.057	.482	.482	1905 (73%)
$r_2/r_1 =$ 1	5,328	<.001	.034	.050	.078	.712	.712	3,288 (62%)	<.001	.035	.050	.057	.484	.484	3,950 (74%)
2	7,776	<.001	.033	.051	.080	.719	.719	4,743 (61%)	<.001	.034	.050	.057	.484	.484	5,720 (74%)
4	5,184	<.001	.034	.050	.082	.724	.724	3,196 (62%)	<.001	.034	.050	.057	.483	.483	3814 (74%)
8	2,592	<.001	.035	.050	.084	.725	.725	1,603 (62%)	<.001	.035	.050	.057	.482	.482	1,905 (73%)

[1] Range = Max - Min

[2] Number (%) of cases with $\alpha \in (.025, .10)$

Table 2.5: Type I error by $m_2/m_1 \times \bar{n}_2/\bar{n}_1$

			Means Unweighted								Means Weighted by Cluster Size						
		N	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]		Min	Q1	Med	Q3	Max	R[1]	N (%)[2]	
$\frac{m_2}{m_1} = 1$	$\frac{\bar{n}_2}{\bar{n}_1} = 1$	720	.038	.049	.050	.050	.050	.012	720	(100%)	.050	.050	.052	.056	.091	.041	720	(100%)
1	2	960	.040	.050	.050	.050	.053	.013	960	(100%)	.039	.048	.050	.050	.080	.041	960	(100%)
1	4	768	.045	.050	.051	.052	.060	.015	768	(100%)	.017	.031	.042	.048	.056	.038	691	(90%)
1	8	576	.049	.051	.052	.055	.068	.019	576	(100%)	.004	.016	.030	.044	.049	.045	347	(60%)
1	16	384	.050	.051	.053	.058	.076	.026	384	(100%)	<.001	.007	.020	.041	.047	.046	146	(38%)
1	32	192	.050	.052	.054	.060	.082	.031	192	(100%)	<.001	.003	.010	.039	.043	.043	64	(33%)
$\frac{m_2}{m_1} = 2$	$\frac{\bar{n}_2}{\bar{n}_1} = 1/32$	192	.008	.012	.017	.028	.035	.027	65	(34%)	.002	.018	.034	.045	.068	.066	112	(58%)
2	1/16	384	.009	.014	.019	.030	.041	.032	140	(36%)	.009	.029	.044	.048	.099	.090	316	(82%)
2	1/8	576	.011	.017	.022	.033	.045	.034	244	(42%)	.028	.045	.049	.051	.132	.103	566	(98%)
2	1/4	768	.016	.022	.028	.038	.049	.033	469	(61%)	.050	.051	.055	.064	.153	.103	748	(97%)
2	1/2	960	.025	.032	.037	.044	.051	.026	959	(100%)	.050	.052	.058	.069	.148	.098	936	(98%)
2	1	1,152	.036	.048	.050	.051	.060	.024	1,152	(100%)	.050	.050	.052	.057	.119	.069	1146	(99%)
2	2	960	.050	.057	.066	.075	.091	.041	960	(100%)	.023	.037	.046	.049	.076	.053	949	(99%)
2	4	768	.052	.068	.088	.106	.126	.073	484	(63%)	.005	.018	.033	.045	.050	.044	480	(63%)
2	8	576	.056	.076	.115	.136	.166	.110	221	(38%)	.001	.008	.021	.041	.048	.048	274	(48%)
2	16	384	.063	.081	.137	.159	.213	.149	128	(33%)	<.001	.003	.013	.037	.046	.046	133	(35%)
2	32	192	.077	.087	.158	.176	.250	.172	64	(33%)	<.001	.001	.005	.036	.041	.041	64	(33%)
$\frac{m_2}{m_1} = 4$	$\frac{\bar{n}_2}{\bar{n}_1} = 1/32$	144	<.001	.001	.003	.016	.022	.021	0	(0%)	.033	.046	.050	.056	.165	.132	135	(94%)
4	1/16	288	.001	.002	.004	.018	.032	.031	43	(15%)	.050	.052	.061	.083	.221	.171	243	(84%)
4	1/8	432	.001	.004	.008	.022	.041	.039	99	(23%)	.051	.054	.073	.110	.251	.200	294	(68%)
4	1/4	576	.004	.009	.016	.030	.047	.043	178	(31%)	.051	.056	.075	.115	.239	.188	376	(65%)
4	1/2	720	.013	.022	.032	.042	.053	.040	485	(67%)	.050	.054	.069	.094	.187	.136	587	(82%)
4	1	864	.036	.049	.051	.059	.096	.060	864	100%	.050	.050	.052	.058	.114	.064	855	(99%)
4	2	720	.051	.067	.089	.108	.175	.124	457	(63%)	.015	.029	.042	.048	.050	.035	602	(84%)
4	4	576	.055	.089	.145	.178	.268	.212	182	(32%)	.002	.010	.027	.043	.049	.047	294	(51%)
4	8	432	.063	.103	.200	.247	.352	.288	97	(22%)	<.001	.002	.013	.037	.047	.047	158	(37%)
4	16	288	.079	.115	.251	.303	.416	.337	40	(14%)	<.001	<.001	.008	.034	.043	.043	96	(33%)
4	32	144	.109	.132	.291	.341	.459	.350	0	(0%)	<.001	<.001	.003	.033	.036	.036	48	(33%)

[1] Range = Max - Min

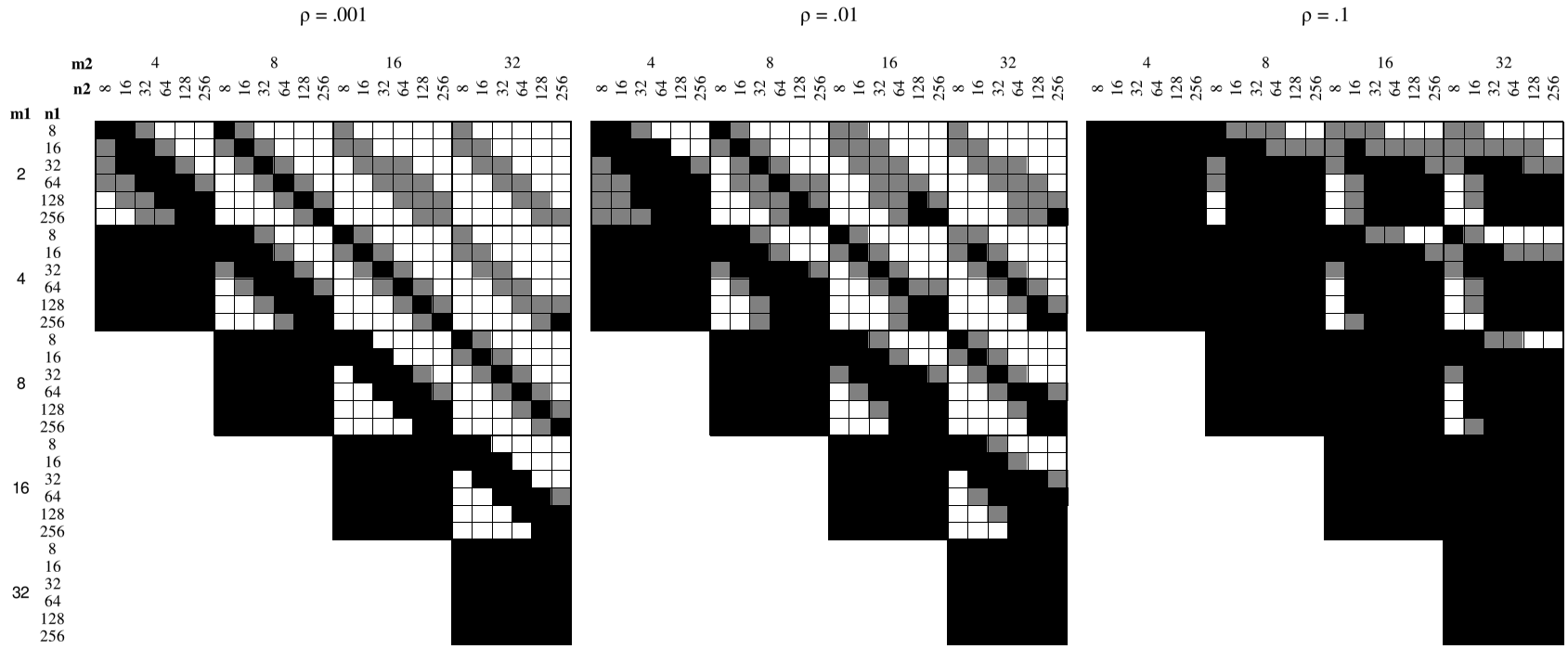
[2] Number (%) of cases with $\alpha \in (.025, .10)$

Table 2.6: Type I error by $m_2/m_1 \times \bar{n}_2/\bar{n}_1$ (cont.)

			Means Unweighted							Means Weighted by Cluster Size						
		N	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]	Min	Q1	Med	Q3	Max	R[1]	N (%)[2]
$\frac{m_2}{m_1} = 8$	$\frac{\bar{n}_2}{\bar{n}_1} = 1/32$	96	<.001	<.001	<.001	.010	.014	.014	0 (0%)	.053	.055	.087	.175	.330	.277	56 (58%)
8	1/16	192	<.001	<.001	.001	.012	.026	.026	2 (1%)	.053	.058	.103	.208	.374	.320	92 (48%)
8	1/8	288	<.001	.001	.003	.016	.038	.037	43 (15%)	.052	.060	.102	.204	.368	.316	142 (49%)
8	1/4	384	.001	.004	.011	.026	.046	.045	101 (26%)	.051	.061	.096	.174	.312	.261	199 (52%)
8	1/2	480	.008	.017	.030	.043	.060	.052	296 (62%)	.051	.056	.077	.117	.218	.168	306 (64%)
8	1	576	.035	.049	.053	.070	.150	.115	523 (91%)	.050	.050	.053	.062	.115	.065	563 (98%)
8	2	480	.052	.075	.110	.142	.276	.224	201 (42%)	.011	.025	.040	.047	.050	.039	358 (75%)
8	4	384	.058	.105	.191	.246	.405	.348	86 (22%)	.001	.007	.023	.041	.048	.047	181 (47%)
8	8	288	.069	.127	.276	.351	.512	.443	44 (15%)	<.001	.001	.011	.034	.046	.046	96 (33%)
8	16	192	.092	.151	.356	.436	.590	.498	16 (8%)	<.001	<.001	.005	.032	.041	.041	64 (33%)
8	32	96	.135	.180	.408	.504	.641	.506	0 (0%)	<.001	<.001	.002	.031	.033	.033	32 (33%)
$\frac{m_2}{m_1} = 16$	$\frac{\bar{n}_2}{\bar{n}_1} = 1/32$	48	<.001	<.001	<.001	.007	.010	.010	0 (0%)	.058	.063	.145	.348	.474	.415	16 (33%)
16	1/16	96	<.001	<.001	<.001	.009	.022	.022	0 (0%)	.056	.066	.151	.341	.484	.428	32 (33%)
16	1/8	144	<.001	<.001	.002	.014	.036	.036	18 (13%)	.054	.066	.137	.299	.445	.391	56 (39%)
16	1/4	192	<.001	.003	.012	.025	.046	.045	47 (24%)	.052	.065	.115	.221	.359	.307	88 (46%)
16	1/2	240	.006	.017	.031	.049	.069	.064	150 (63%)	.051	.058	.084	.137	.239	.188	136 (57%)
16	1	288	.035	.050	.057	.088	.183	.148	236 (82%)	.050	.050	.054	.066	.116	.066	272 (94%)
16	2	240	.053	.082	.127	.181	.330	.278	84 (35%)	.009	.024	.038	.046	.050	.041	176 (73%)
16	4	192	.059	.117	.233	.302	.472	.413	36 (19%)	.001	.005	.021	.040	.048	.047	88 (46%)
16	8	144	.073	.156	.341	.426	.587	.514	16 (11%)	<.001	<.001	.009	.034	.045	.045	48 (33%)
16	16	96	.100	.182	.419	.527	.670	.570	0 (0%)	<.001	<.001	.004	.031	.040	.040	32 (33%)
16	32	48	.151	.233	.503	.611	.725	.574	0 (0%)	<.001	<.001	.001	.030	.031	.031	16 (33%)

[1] Range = Max - Min
[2] Number (%) of cases with $\alpha \in (.025, .10)$

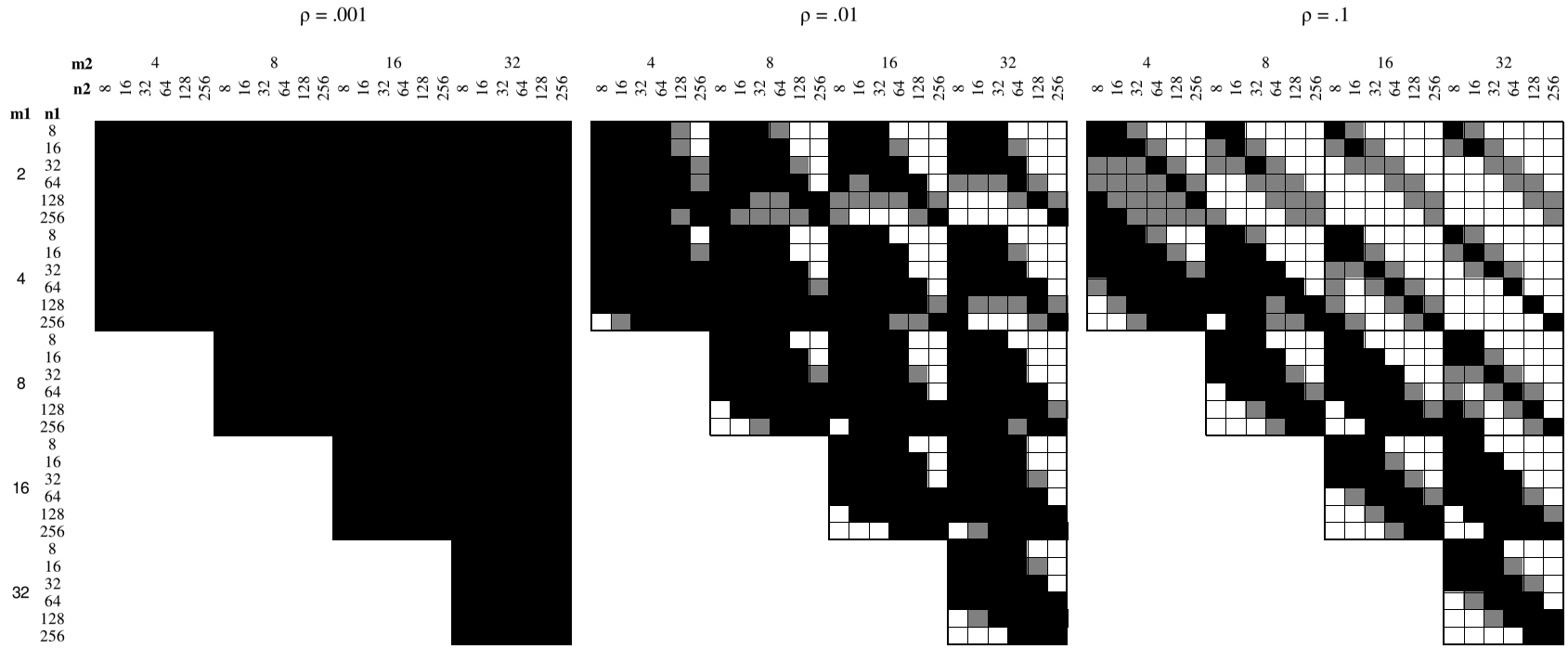
Figure 2.1: Type I Error for $\rho \times m_1 \times m_2 \times \bar{n}_1 \times \bar{n}_2$ - Analysis of Unweighted Means



Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) combinations of $r_1 \times r_2$.

Black: All combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
Grey: Some combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
White : No combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$

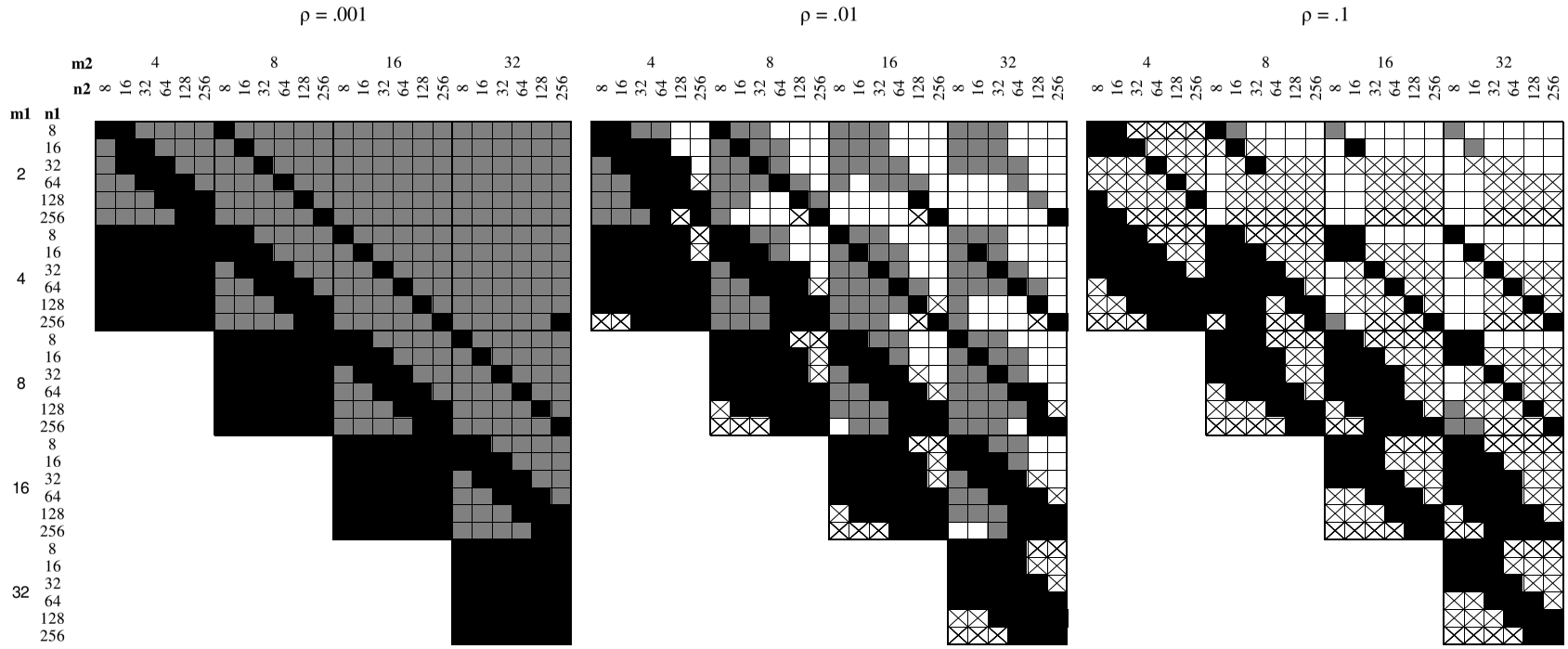
Figure 2.2: Type I Error for $\rho \times m_1 \times m_2 \times \bar{n}_1 \times \bar{n}_2$ - Analysis of Means Weighted By Cluster Size



Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) combinations of $r_1 \times r_2$.

Black: All combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
Grey: Some combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
White : No combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$

Figure 2.3: Type I Error for $\rho \times m_1 \times m_2 \times \bar{n}_1 \times \bar{n}_2$ - Comparison of Weights



Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) combinations of $r_1 \times r_2$.

Black: For both weightings, all combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
(Both weightings are unbiased)

Grey: For weighting means by cluster size, all combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
(Only cluster size weights are unbiased)

Cross: For unweighted means, all combinations of $r_1 \times r_2$ have hypothesis tests with $\alpha \in (.025, .10)$
(Only unweighted are unbiased)

White: Neither weighting scheme has all combinations of $r_1 \times r_2$ with hypothesis tests with $\alpha \in (.025, .10)$
(Neither weighting unbiased)

Chapter 3

Comparison of Type I Error for One Stage and Two Stage Models

3.1 Introduction

Cluster or group randomized trials are those in which data are collected on (correlated) individuals within specific geographic or civil units (clusters) and clusters, not individuals, are randomized to the explanatory variable of interest [7, 24]. Throughout this paper we will refer to the explanatory (and randomization) variable of interest as treatment group (the usual choice in group randomized trials), though discussions here apply to any cluster level explanatory variable. Murray [24] and Donner and Klar [7] provided a thorough history of the use of cluster randomized trials and give examples which span the wide variety of areas of medicine and public health in which cluster randomized trials have been used.

Murray [24] described methods commonly used for the analysis of clustered data, in particular, analysis via a "one stage" or "two stage" linear model. Though such analyses have been performed on categorical outcome data with various error distributions, we focus on the analysis of Gaussian clustered data in this paper.

In the one stage analysis, correlated individual level data are analyzed directly via a mixed effects linear model. In the two stage analysis, cluster means are computed first, often adjusted for covariates through a preliminary model excluding treatment group, and cluster means are the values of the response in a linear model at the second stage. Varnell *et al.* [35] showed that current researchers use both approaches. They reviewed group randomized trials published in the *American Journal of Public Health* and *Preventive Medicine* from 1998 to 2002 and showed

that of 47 trials that employed at least one statistical analysis appropriate for group randomized trials, 32 (68%) analyzed individual level data and 15 (32%) analyzed cluster means or another summary statistic.

If data have a common within-cluster correlation, a common individual error variance, and equal numbers of observations per cluster both the two-stage and one-stage approaches give the same test statistic. This special feature is due to the special compound symmetric (equal correlation) covariance structure of clustered data. Further, this test is the uniformly most powerful size-α test and has exact null and non-null distributions. Also, this test statistic is derived from closed formed expressions for the maximum likelihood estimates for the fixed effects and variance components, which have known distributions.

For data with one level of clustering, if any of the previous conditions about common variance, correlation, or cluster size do not hold, the one-stage and two-stage analysis approaches lead to different tests. No uniformly most powerful size-α test for the fixed effects exist; the unbalanced versions of the test statistics used for balanced data now have only approximate distributions; and closed form expressions for estimates of variance components are no longer available. Research is needed to study the distributional properties of the hypothesis test statistics for fixed effects in the one-stage and two-stage analysis of unbalanced clustered data. In this paper, we conduct simulations of type I error for these tests with Gaussian data for scenarios of imbalance in cluster sizes commonly found in cluster randomized trials.

3.2 Statement of Models and Hypothesis

3.2.1 Matrix Notation

Lower case bold, upper case bold, and upper case italics indicate a (column) vector, a matrix, and a random variable, respectively.

We use the following notation of McCulloch and Searle [19, Appendix M, Section 3], to conveniently denote stacked column vectors and diagonal matrices with similar notation. Define indices i and j such that $i = 1, ..., a$ and $j = 1, ..., b$. Let $\boldsymbol{u} = \left\{ {}_c\, \boldsymbol{u}_{ij} \right\}$ denote the stacked column vector $\boldsymbol{u}$, where $\boldsymbol{u} = \left\{ \boldsymbol{u}'_{11} \quad \boldsymbol{u}'_{12} \quad \dots \quad \boldsymbol{u}'_{ij} \quad \dots \quad \boldsymbol{u}'_{ab} \right\}'$. Let $\boldsymbol{U} = \left\{ {}_d\, \boldsymbol{U}_{ij} \right\}$ denote the diagonal matrix $\boldsymbol{U}$ with diagonal elements $\boldsymbol{U}_{11}, ..., \boldsymbol{U}_{ab}$, i.e. $\boldsymbol{U} = \mathrm{Dg}\left(\boldsymbol{U}_{11}, \boldsymbol{U}_{12}, ..., \boldsymbol{U}_{ij}, ..., \boldsymbol{U}_{ab}\right)$.

Kronecker product multiplication of matrix $\boldsymbol{A}$ by matrix $\boldsymbol{B}$ is denoted by $\boldsymbol{A} \otimes \boldsymbol{B} = \{a_{ij}\boldsymbol{B}\}$

where a_{ij} is the element in the i-th row and j-th column of matrix $\boldsymbol{A}$. All other matrix operators are defined as in standard practice; Muller and Stewart [23] or Schott [30] gave details.

3.2.2 Data Structure

We consider data with one level of clustering, so that individual observations are nested within cluster. In the following notation, clusters are also nested within treatment group. Let $h = 1, \ldots, g$ index treatment groups, $i = 1, \ldots, m_h$ index clusters within treatment group h, and let $j = 1, \ldots, n_{hi}$ index observations within treatment cluster i and treatment group h. Denote the total number of clusters by $m = \sum_{h=1}^{g} m_h$, the number of observations in treatment group h by $n_h = \sum_{i=1}^{m_h} n_{hi}$, and the total number of observations by $N = \sum_{h=1}^{g} \sum_{i=1}^{m_h} n_{hi}$.

3.2.3 Statement of One-Stage Model

Define a linear model for continuous Gaussian outcome $\boldsymbol{y_1}$ that includes fixed effects given in $\boldsymbol{\beta}$ $(g \times 1)$, a random effect for cluster given in $\boldsymbol{b}$ $(m \times 1)$, and a random error, $\boldsymbol{e}_1$ $(N \times 1)$:

$$\boldsymbol{y}_1 = \boldsymbol{X}_1\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{b} + \boldsymbol{e}_1. \tag{3.1}$$

The matrices $\boldsymbol{X}_1$ $(N \times g)$ and $\boldsymbol{Z}_1$ $(N \times m)$ are design matrices for the fixed and random effects, respectively. The simulations considered here did not include any fixed covariates other than treatment group nor any random effects other than cluster, so that $\boldsymbol{X}_1$ contains only an effect for treatment group and $\boldsymbol{Z}_1$ only an effect for cluster.

Vectors or matrices $\boldsymbol{y}_1$, $\boldsymbol{X}_1$, $\boldsymbol{Z}_1$,and $\boldsymbol{e}_1$ are stacked by treatment group and cluster so that $\boldsymbol{y}_1 = \left\{_c\, \boldsymbol{y}_{1,hi}\right\}$, $\boldsymbol{X}_1 = \left\{_c\, \boldsymbol{X}_{1,hi}\right\}$, $\boldsymbol{Z}_1 = \left\{_c\, \boldsymbol{Z}_{1,hi}\right\}$, and $\boldsymbol{e}_1 = \left\{_c\, \boldsymbol{e}_{1,hi}\right\}$. Without loss of generality, assume the fixed effects design matrix $\boldsymbol{X}_1$ has a cell mean coding for treatment group so that $\boldsymbol{X}_1 = \left\{_d\, \mathbf{1}_{n_h}\right\}$. The design matrix for the random cluster effect is $\boldsymbol{Z}_1 = \left\{_d\, \mathbf{1}_{n_{hi}}\right\}$.

We assume $\boldsymbol{b} \sim \mathcal{N}_m\left(\mathbf{0}, \sigma_c^2 \boldsymbol{I}_m\right)$ independently of $\boldsymbol{e}_1 \sim \mathcal{N}_N\left(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_N\right)$ so that:

$$\boldsymbol{y}_1 \sim \mathcal{N}_N\left(\boldsymbol{X}_1\boldsymbol{\beta}, \boldsymbol{\Sigma}_1\right),$$

where the covariance matrix $\boldsymbol{\Sigma}_1$ $(N \times N)$ is compound symmetric and has the form:

$$\boldsymbol{\Sigma}_1 = \sigma_c^2 \boldsymbol{Z}_1\boldsymbol{Z}_1' + \sigma_e^2 \boldsymbol{I}_N = \left\{_d\, \sigma_c^2 \mathbf{1}_{n_{hi}} \mathbf{1}_{n_{hi}}' + \sigma_e^2 \boldsymbol{I}_{n_{hi}}\right\}.$$

$\boldsymbol{\Sigma}_1$ may be expressed in terms of the total variance, σ_y^2, and within cluster correlation, ρ, as:

$$\boldsymbol{\Sigma}_1 = \sigma_y^2 \{_d\, \mathbf{1}_{n_{hi}} \mathbf{1}'_{n_{hi}} \rho + \boldsymbol{I}_{n_{hi}} (1-\rho)\},$$

where $\rho = \sigma_c^2 / \left(\sigma_c^2 + \sigma_e^2\right)$ and $\sigma_y^2 = \sigma_c^2 + \sigma_e^2$ or, equivalently, $\sigma_c^2 = \sigma_y^2 \rho$ and $\sigma_e^2 = \sigma_y^2 (1-\rho)$. Implicit in construction of $\boldsymbol{\Sigma}_1$ is the assumption that data across all treatment groups have the same variance parameters.

Common statistical software, such as PROC MIXED in SAS, estimate variance components in the $\left(\sigma_c^2, \sigma_e^2\right)$ parameterization; however, clustered data studies are more often planned with estimates in the $\left(\sigma_y^2, \rho\right)$ parameterization. We alternate between these when appropriate.

3.2.4 Statement of Two-Stage Model

To transform from a model for individual level data to a model for cluster means, pre-multiply model (3.1) by the matrix $\boldsymbol{T}_1$ $(m \times N)$, where $\boldsymbol{T}_1 = \{_d\, \mathbf{1}'_{n_{hi}} / n_{hi}\}$. This yields a model for $\boldsymbol{y}_2$ $(m \times 1) = \boldsymbol{T}_1 \boldsymbol{y}_1$ where:

$$\boldsymbol{y}_2 = \boldsymbol{X}_2 \boldsymbol{\beta} + \boldsymbol{Z}_2 \boldsymbol{b} + \boldsymbol{e}_2, \tag{3.2}$$

and $\boldsymbol{X}_2$ $(m \times g) = \boldsymbol{T}_1 \boldsymbol{X}_1$, $\boldsymbol{Z}_2$ $(m \times m) = \boldsymbol{T}_1 \boldsymbol{Z}_1$, and $\boldsymbol{e}_2$ $(m \times 1) = \boldsymbol{T}_1 \boldsymbol{e}_1$. Parameters $\boldsymbol{\beta}$ and $\boldsymbol{b}$ were not affected by the transformation.

The vector of outcomes, $\boldsymbol{y}_2$, and of random errors, $\boldsymbol{e}_2$, contain cluster averages, so that $\boldsymbol{y}_2 = \{_c\, \overline{y}_{hi}\}$ and $\boldsymbol{e}_2 = \{_c\, \overline{e}_{hi}\}$. The fixed and random effects design matrices are $\boldsymbol{X}_2 = \{_d\, \mathbf{1}'_{n_{hi}} / n_{hi}\} \{_d\, \mathbf{1}_{n_h}\} = \{_d\, \mathbf{1}_{m_h}\}$ and $\boldsymbol{Z}_2 = \{_d\, \mathbf{1}'_{n_{hi}} / n_{hi}\} \{_d\, \mathbf{1}_{n_{hi}}\} = \boldsymbol{I}_m$.

In line with previous assumptions, we assume $\boldsymbol{b} \sim \mathcal{N}_m \left(\mathbf{0}, \sigma_c^2 \boldsymbol{I}_m\right)$ independently of $\boldsymbol{e}_2 \sim \mathcal{N}_m \left(\mathbf{0}, \sigma_e^2 \boldsymbol{T}_1 \boldsymbol{T}'_1\right)$ so that:

$$\boldsymbol{y}_2 \sim \mathcal{N}_m \left(\boldsymbol{X}_2 \boldsymbol{\beta}, \boldsymbol{\Sigma}_2\right),$$

where $\boldsymbol{\Sigma}_2$ $(m \times m)$ is given by:

$$\boldsymbol{\Sigma}_2 = \boldsymbol{T}_1 \boldsymbol{\Sigma}_1 \boldsymbol{T}'_1 = \{_d\, \sigma_c^2 + \sigma_e^2 / n_{hi}\}.$$

In terms of the alternate parameterization with $\left(\sigma_y^2, \rho\right)$ instead of $\left(\sigma_e^2, \sigma_c^2\right)$:

$$\boldsymbol{\Sigma}_2 = \sigma_y^2 \{_d\, [1 + (n_{hi} - 1)\, \rho] / n_{hi}\}.$$

3.2.5 General Linear Hypothesis

Define a vector of secondary contrast parameters, $\boldsymbol{\theta}\ (a \times 1) = \boldsymbol{C\beta}$, where $\boldsymbol{C}\ (a \times g)$ contains desired contrasts for the fixed effects. We study the two sided general linear hypothesis (GLH) $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. In most hypotheses of interest, $\boldsymbol{\theta}_0 = \mathbf{0}$. Such a hypothesis test describes differences in the fixed effects only. The $\boldsymbol{X}$ matrices we consider are full rank, ensuring $\boldsymbol{\theta}$ is estimable. Requiring $\boldsymbol{C}$ to have full row rank [$\text{rank}(\boldsymbol{C}) = a$] ensures the GLH is testable [21].

3.3 Hypothesis Testing for Clustered Data with Balanced Cluster Sizes

The analysis of clustered data has special properties when data have balanced cluster sizes. Estimation and hypothesis testing of fixed effects for such balanced data are discussed in this section. In practice, these methods are applied to unbalanced clustered data as well. The properties of methods of estimation and hypothesis testing for balanced clustered data when applied to unbalanced data are discussed in Section 3.4.

3.3.1 Estimation of Fixed Effects for the One Stage Model for Individual Data

Consider the one-stage model for individual level data $\boldsymbol{y}_1 \sim \mathcal{N}_N\left(\boldsymbol{X}_1\boldsymbol{\beta}, \boldsymbol{\Sigma}_1\right)$ given in model 3.1. When $\boldsymbol{\Sigma}_1$ is unknown, and therefore contains nuisance parameters which must be estimated, the restricted maximum likelihood (REML) estimator for $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{y}_1,$$

where $\widehat{\boldsymbol{\Sigma}}_1\ (N \times N)$ is the REML estimator for $\boldsymbol{\Sigma}_1$. When individual level clustered data $\boldsymbol{y}_1$ have balanced cluster sizes, that is, when $n_{hi} \equiv n$ for all h,i, the estimator $\widehat{\boldsymbol{\Sigma}}_1$ can be stated in terms of a Kronecker product of the same covariance matrix for all clusters. That is, $\widehat{\boldsymbol{\Sigma}}_1 = \boldsymbol{I}_m \otimes \left\{\widehat{\sigma}_y^2\left[\mathbf{1}_n\mathbf{1}_n'\widehat{\rho} + \boldsymbol{I}_n\left(1-\widehat{\rho}\right)\right]\right\}$. Because of this, the inverse of $\widehat{\boldsymbol{\Sigma}}_1$ can be written with the closed form expression:

$$\widehat{\boldsymbol{\Sigma}}_1^{-1} = \boldsymbol{I}_m \otimes \frac{1}{\widehat{\sigma}_y^2\left(1-\widehat{\rho}\right)}\left\{\boldsymbol{I}_n - \frac{\widehat{\rho}}{\left[1+(n-1)\,\widehat{\rho}\right]}\mathbf{1}_n\mathbf{1}_n'\right\}.$$

For balanced data with no covariates, $\boldsymbol{X}_1 = \left\{_d \mathbf{1}_{m_h n}\right\}$, and we can show that $\boldsymbol{X}_1' \widehat{\boldsymbol{\Sigma}}_1^{-1} = \left\{\widehat{\sigma}_y^2 \left[1 + (n-1)\,\widehat{\rho}\right]\right\}^{-1} \boldsymbol{X}_1'$. Thus, for balanced data:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= \left(\boldsymbol{X}_1' \widehat{\boldsymbol{\Sigma}}_1^{-1} \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1' \widehat{\boldsymbol{\Sigma}}_1^{-1} \boldsymbol{y}_1 \\
&= \left(\left\{\widehat{\sigma}_y^2 \left[1 + (n-1)\,\widehat{\rho}\right]\right\}^{-1} \boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} \left\{\widehat{\sigma}_y^2 \left[1 + (n-1)\,\widehat{\rho}\right]\right\}^{-1} \boldsymbol{X}_1' \boldsymbol{y}_1 \\
&= \left(\boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1' \boldsymbol{y}_1.
\end{aligned}$$

We now have an estimator for the fixed effects that does not depend on the variance components. That is, for balanced data, the weighted least squares and ordinary least squares estimators of $\boldsymbol{\beta}$ coincide. Puntanen and Styan [26] and Tian and Wiens [34] gave a comprehensive review of when weighted or ordinary least squares estimators coincide for general data $\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$.

3.3.2 Estimation of Fixed Effects for the Two Stage Model for Cluster Means

Consider the two-stage model for cluster means $\boldsymbol{y}_2 \sim \mathcal{N}_m\left(\boldsymbol{X}_2 \boldsymbol{\beta}, \boldsymbol{\Sigma}_2\right)$ given in model 3.2. When data are balanced, $n_{hi} \equiv n$ for all h, i, so that the cluster means have covariance:

$$\boldsymbol{\Sigma}_2 = \boldsymbol{I}_m \otimes \left(\sigma_y^2 / n\right)\left\{1 + (n-1)\,\rho\right\}.$$

That is, for balanced data, all cluster means have the same variance, and $\boldsymbol{\Sigma}_2$ can be written as $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{I}_m$ where: $\sigma^2 = \left(\sigma_y^2 / n\right)\left\{1 + (n-1)\,\rho\right\}$. Independence, normality, and homogeneity of errors of the cluster means meet the assumptions of the general linear univariate model (GLUM).

In the general linear univariate model, the best linear unbiased and maximum likelihood estimator for the fixed effects, $\boldsymbol{\beta}$, is:

$$\widehat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2' \boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2' \boldsymbol{y}_2.$$

3.3.3 Equivalence of Estimators from One and Two Stage Models

Matrix algebra shows that:

$$\begin{aligned}
\boldsymbol{X}_1' \boldsymbol{X}_1 &= \left\{_d \mathbf{1}_{m_h n}'\right\}\left\{_d \mathbf{1}_{m_h n}\right\} = \left\{_d n m_h\right\} \\
\boldsymbol{X}_1' \boldsymbol{y}_1 &= \left\{_d \mathbf{1}_{m_h n}'\right\}\left\{_c \boldsymbol{y}_{1,hi}\right\} = \left\{_c \sum_{i=1}^{m_h} n \overline{y}_{hi}\right\}
\end{aligned}$$

as well as:

$$\boldsymbol{X}_2'\boldsymbol{X}_2 = \left\{{}_d\,\mathbf{1}_{m_h}'\right\}\left\{{}_d\,\mathbf{1}_{m_h}\right\} = \left\{{}_d\,m_h\right\}$$
$$\boldsymbol{X}_2'\boldsymbol{y}_2 = \left\{{}_d\,\mathbf{1}_{m_h}'\right\}\left\{{}_c\,\overline{y}_{hi}\right\} = \left\{{}_c\sum_{i=1}^{m_h}\overline{y}_{hi}\right\}.$$

Thus:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1 = \left\{{}_d\,nm_h\right\}\left\{{}_c\sum_{i=1}^{m_h}n\overline{y}_{hi}\right\} = \left\{{}_c\,\overline{y}_h\right\}$$

and:

$$\widehat{\boldsymbol{\beta}}_2 = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{y}_2 = \left\{{}_d\,m_h\right\}\left\{{}_c\sum_{i=1}^{m_h}\overline{y}_{hi}\right\} = \left\{{}_c\,\overline{y}_h\right\}.$$

That is, for data with balanced cluster sizes, the population treatment means in $\boldsymbol{\beta}$ are estimated by the sample treatment means in both the one-stage and two-stage models.

3.3.4 Hypothesis Test for Fixed Effects

The equivalent estimators given in sections 3.3.1 and 3.3.2 may be written as

$$\widehat{\boldsymbol{\beta}}_s = \left(\boldsymbol{X}_s'\boldsymbol{X}_s\right)^{-1}\boldsymbol{X}_s'\boldsymbol{y}_s$$

where $s = 1, 2$. Using theory of quadratic forms in normal vectors, $\widehat{\boldsymbol{\beta}}_s \sim \mathcal{N}_g\left[\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}_s'\boldsymbol{X}_s\right)^{-1}\right]$, where $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$. Further, an estimator for desired contrasts, $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$, may be estimated with $\widehat{\boldsymbol{\theta}}_s = \boldsymbol{C}\widehat{\boldsymbol{\beta}}_s$, which has distribution $\widehat{\boldsymbol{\theta}}_s \sim \mathcal{N}_a\left[\boldsymbol{\theta}, \sigma^2\boldsymbol{C}\left(\boldsymbol{X}_s'\boldsymbol{X}_s\right)^{-1}\boldsymbol{C}'\right]$.

Using theory for a general linear univariate model, a uniformly most powerful size α test for the GLH is given by:

$$T_s = \left(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_0\right)'\left[\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_s\right)\right]^{-1}\left(\widehat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_0\right)/a,$$

where $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_s\right)$ is the estimated variance of $\widehat{\boldsymbol{\theta}}_s$, that is $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_s\right) = \widehat{\sigma}^2\left[\boldsymbol{C}\left(\boldsymbol{X}_s'\boldsymbol{X}_s\right)^{-1}\boldsymbol{C}'\right]$. The quantity $\widehat{\sigma}^2$ denotes the restricted maximum likelihood estimator for σ^2, discussed in the next section. This test statistic can be shown to have distribution:

$$T_s \sim \mathcal{F}\left(a, m-g, \omega\right),$$

where $\omega = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)'\left[\boldsymbol{C}\left(\boldsymbol{X}_s'\boldsymbol{X}_s\right)^{-1}\boldsymbol{C}'\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)/\sigma^2$.

3.3.4.1 Estimation of Variance Components

In the two stage analysis of cluster means, variance components $\left(\sigma_y^2, \rho\right)$ or $\left(\sigma_c^2, \sigma_e^2\right)$ are not separately estimable so that the linear combination $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$ is estimated.

An estimator for σ^2 in the two stage analysis of cluster means is:

$$\widehat{\sigma}^2 = \boldsymbol{y}_2' \left[\boldsymbol{I}_m - \boldsymbol{X}_2 \left(\boldsymbol{X}_2' \boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_2' \right] \boldsymbol{y}_2 / \left(m - g \right).$$

This estimator can be derived as a maximum likelihood and ANOVA estimator:

$$\widehat{\sigma}^2 = \mathrm{SSE}_2 / \left(m - g \right),$$

where SSE_2 denotes the sums of squares error.

In the one stage analysis of individual level data, though an estimator of the linear combination $\sigma^2 = \left(\sigma_y^2 / n \right) \left\{ 1 + (n-1)\, \rho \right\} = \sigma_c^2 + \sigma_e^2 / n$ is needed, current statistical software, designed for the estimation of parameters of general covariance structures, estimates the variance components separately in the parameterization $\left(\sigma_c^2, \sigma_e^2 \right)$.

Define the sums of squares due to cluster and error, respectively, as:

$$\mathrm{SSC}_1 = \boldsymbol{y}_1' \left[\boldsymbol{Z}_1 \left(\boldsymbol{Z}_1' \boldsymbol{Z}_1 \right)^{-1} \boldsymbol{Z}_1 - \boldsymbol{X}_1 \left(\boldsymbol{X}_1' \boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1 \right] \boldsymbol{y}_1$$

$$\mathrm{SSE}_1 = \boldsymbol{y}_1' \left[\boldsymbol{I}_N - \boldsymbol{Z}_1 \left(\boldsymbol{Z}_1' \boldsymbol{Z}_1 \right)^{-1} \boldsymbol{Z}_1' \right] \boldsymbol{y}_1.$$

as well as mean squares due to cluster and error, respectively:

$$\mathrm{MSC}_1 = \mathrm{SSC}_1 / \left(m - g \right)$$

$$\mathrm{MSE}_1 = \mathrm{SSE}_1 / \left(N - m \right).$$

The parameterization $\left(\sigma_c^2, \sigma_e^2 \right)$ requires estimates of both σ_c^2 and σ_c^2 to be positive, since σ_c^2 and σ_e^2 are defined as variances. As such, several sources, e.g. Searle [31, p .419], point out that restricted maximum likelihood estimators for σ_c^2 and σ_e^2 are given by:

$$\widehat{\sigma}_c^2 = \left(\mathrm{MSC}_1 - \mathrm{MSE}_1 \right) / n$$

$$\widehat{\sigma}_e^2 = \mathrm{MSE}_1$$

when $\mathrm{MSC}_1 \geq \mathrm{MSE}_1$ (that is, when $\widehat{\sigma}_c^2$ is positive) and

$$\widehat{\sigma}_c^2 = 0$$

$$\widehat{\sigma}_e^2 = \mathrm{SST}_1 / \left(N - m \right)$$

when $\mathrm{MSC}_1 < \mathrm{MSE}_1$, where $\mathrm{SST}_1 = \mathrm{SSC}_1 + \mathrm{SSE}_1$ denotes the total sums of squares of the individual level data. The probability that $\widehat{\sigma}_c^2 < 0$ is:

$$\Pr \left\{ \widehat{\sigma}_c^2 < 0 \right\} = \Pr \left\{ \mathcal{F}_{m-1, N-m} < 1 / \left[1 + n\rho / \left(1 - \rho \right) \right] \right\}.$$

It can be shown that when $\mathrm{MSC}_1 \geq \mathrm{MSE}_1$, the estimator $\widehat{\sigma}^2 = \widehat{\sigma}_c^2 + \widehat{\sigma}_e^2 / n$ is equivalent to

the best linear unbiased and maximum likelihood estimator obtained in the two-stage analysis; however, this is not the case when $\text{MSC}_1 < \text{MSE}_1$. That is, when $\text{MSC}_1 < \text{MSE}_1$, the linear combination of restricted maximum likelihood estimators for each of σ_c^2 and σ_e^2 is NOT the restricted maximum likelihood estimator for the linear combination σ^2. The restricted maximum likelihood estimator for σ^2 is obtained only when variance components estimators are $\widehat{\sigma}_c^2 = (\text{MSC}_1 - \text{MSE}_1)/n$ and $\widehat{\sigma}_e^2 = \text{MSE}_1$, and the estimator $\widehat{\sigma}_c^2$ is allowed to be negative.

Default behavior of SAS PROC MIXED is to constrain estimates of variance components to be positive; simulations in this paper explore the ramifications of this choice.

3.4 Hypothesis Testing for Clustered Data with Unbalanced Cluster Sizes

3.4.1 One Stage Model for Individual Data

When cluster sizes are unbalanced, the weighted least squares estimator $\widehat{\boldsymbol{\beta}}_1$, given in section 3.3.1 as:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{y}_1,$$

is no longer equivalent to an ordinary least squares estimator. That is, estimation of fixed effects now requires estimation of the variance components.

Recall that the structure of $\boldsymbol{\Sigma}_1$ for unbalanced data is:

$$\boldsymbol{\Sigma}_1 = \left\{{}_d\,\sigma_c^2\mathbf{1}_{n_{hi}}\mathbf{1}_{n_{hi}}' + \sigma_e^2\boldsymbol{I}_{n_{hi}}\right\} = \sigma_y^2\left\{{}_d\,\mathbf{1}_{n_{hi}}\mathbf{1}_{n_{hi}}'\rho + \boldsymbol{I}_{n_{hi}}(1-\rho)\right\}.$$

When data are unbalanced, no closed form expressions exist for estimates of the variance components in either parameterization; estimates must be obtained by an iterative procedure such as Newton-Raphson iteration or the EM algorithm [5].

Construction of a hypothesis test for $\boldsymbol{\theta} = \boldsymbol{C}\boldsymbol{\beta}$ requires knowledge of the distribution of $\widehat{\boldsymbol{\beta}}_1$. The estimator $\widehat{\boldsymbol{\beta}}_1$ is unbiased, so that $\mathcal{E}\left(\widehat{\boldsymbol{\beta}}_1\right) = \boldsymbol{\beta}$; however, no closed form expression exists for its variance. The common strategy is to approximately estimate this as $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\beta}}_1\right) = \left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}$. Kacker and Harville [12] and Dempster *et al.* [6] pointed out that this underestimates the true variability in $\widehat{\boldsymbol{\beta}}_1$, since $\left(\boldsymbol{X}_1'\widehat{\boldsymbol{\Sigma}}_1^{-1}\boldsymbol{X}_1\right)^{-1}$ is an estimate of the variance of $\widetilde{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{y}_1$ not of the variance of $\widehat{\boldsymbol{\beta}}_1$.

As in hypothesis testing with balanced data, a hypothesis test for the general linear hypoth-

esis is given by the Wald style test statistic:

$$T_1 = \left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\right)' \left[\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_1\right)\right]^{-1} \left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\right)/a.$$

Since no closed form expressions exist for the estimating elements of $\widehat{\boldsymbol{\Sigma}}_1$, T_1 cannot be written explicitly as a quadratic form, and thus its exact distribution is unknown. Applying large sample theory gives $T_1 \xrightarrow{D} \chi^2(a, \omega)$. Given a random variable X with $X \sim \mathcal{F}(\nu_1, \nu_2, \omega)$, as $\nu_2 \longrightarrow \infty$, $X \xrightarrow{D} Y$ where $Y \sim \chi^2(\nu_1, \omega)$. For this reason, T_1 is usually given an approximate $F(a, \nu_2, \omega)$ distribution in order to combat the underestimate of the variance of $\widehat{\boldsymbol{\beta}}_1$. Several methods exist to estimate the denominator degrees of freedom of T_1. In most cases, none can be shown to be superior [18].

One such method is that proposed by Kenward and Roger [13], who multiply T_1 by an inflation factor to account for the additional variability in $\mathcal{V}\left(\widehat{\boldsymbol{\beta}}_1\right)$ introduced by estimating $\boldsymbol{\Sigma}_1$. Satterthwaite [28] style degrees of freedom are then computed for this inflated statistic. This approximation has not been studied thoroughly in small clustered data settings, and has been shown to be biased in settings with other types of small sample data [3, 15, 29]. Another common choice for denominator degrees of freedom is that for the analysis of balanced data: $\text{ddf} = m - g$.

3.4.2 Two Stage Model for Cluster Means

From section 3.3.2, cluster means with balanced cluster sizes may be analyzed via a general linear univariate model. The general linear univariate model assumes $\boldsymbol{\Sigma}_2$ is proportional to an identity matrix, that is $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{I}$, where as before $\sigma^2 = \left(\sigma_y^2/n\right)\left\{1 + (n-1)\,\rho\right\}$.

An optimal test for the general linear hypothesis can be derived with the less restrictive assumption that $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{W}^{-1}$, that is, that the covariance matrix is known up to a constant weight matrix $\boldsymbol{W}$ [19]. In this case, the best linear unbiased and maximum likelihood estimator for the fixed effects is:

$$\widehat{\boldsymbol{\beta}}_{2w} = \left(\boldsymbol{X}_2' \boldsymbol{W} \boldsymbol{X}_2\right)^{-1} \boldsymbol{X}_2' \boldsymbol{W} \boldsymbol{y}_2.$$

With $\widehat{\boldsymbol{\beta}}_{2w} \sim \mathcal{N}_g\left[\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}_2' \boldsymbol{W} \boldsymbol{X}_2\right)^{-1}\right]$ and thus, $\widehat{\boldsymbol{\theta}}_{2w} = \boldsymbol{C}\widehat{\boldsymbol{\beta}}_{2w} \sim \mathcal{N}_a\left[\boldsymbol{\theta}, \sigma^2 \boldsymbol{C}\left(\boldsymbol{X}_2' \boldsymbol{W} \boldsymbol{X}_2\right)^{-1} \boldsymbol{C}'\right]$, a uniformly most powerful size-α test for the fixed effects can be shown to be given by:

$$T_{2w} = \left(\widehat{\boldsymbol{\theta}}_{2w} - \boldsymbol{\theta}_0\right)' \left[\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_{2w}\right)\right]^{-1} \left(\widehat{\boldsymbol{\theta}}_{2w} - \boldsymbol{\theta}_0\right)/a,$$

where $\widehat{\mathcal{V}}\left(\widehat{\boldsymbol{\theta}}_{2w}\right) = \widehat{\sigma}_w^2 \boldsymbol{C}\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{C}'$. The restricted maximum likelihood estimator for σ^2 is:

$$\widehat{\sigma}_w^2 = \boldsymbol{y}_2'\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}_2\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{W}\right]\boldsymbol{y}_2/\left(m-g\right).$$

The statistic T_{2w} has distribution $T_{2w} \sim \mathcal{F}\left(a,\, m-r,\, \omega_w\right)$, where the noncentrality in the weighted model is $\omega_w = \left(\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{\theta}_0\right)'\left[\boldsymbol{C}\left(\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{X}_2\right)^{-1}\boldsymbol{C}'\right]^{-1}\left(\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{\theta}_0\right)/\sigma^2$.

The exact distribution of the test statistic T_w depends on the assumption that $\boldsymbol{\Sigma}_2$ can be written in the form $\boldsymbol{\Sigma}_2 = \sigma^2\boldsymbol{W}^{-1}$ where $\boldsymbol{W}$ is known. When this doesn't hold T_{2w} has only an approximate $\mathcal{F}$ distribution. Cluster means derived from unbalanced clustered data with $\boldsymbol{\Sigma}_2 = \left(\sigma_y^2\right)\left\{_d\left[1 + \left(n_{hi}-1\right)\rho\right]/n_{hi}\right\}$ do not have this form since variance components must be estimated; however, many types of weights have $\boldsymbol{W}$ such that $\boldsymbol{\Sigma}_2 \approx \sigma^2\boldsymbol{W}^{-1}$.

In an attempt to estimate the variance components in the weights, data analysts also choose $\boldsymbol{W} = \left\{_d[n_{hi}\left(n_{hi}-1\right)]/\left[\boldsymbol{y}_{hi}'\boldsymbol{y}_{hi} - n_{hi}\bar{y}_{1,hi}\right]\right\}$, the diagonal matrix of inverses of the estimated sample variance of each cluster mean, or $\boldsymbol{W} = \left[\left(\widehat{\sigma}_y^2\right)\left\{_d\left[1 + \left(n_{hi}-1\right)\widehat{\rho}\right]/n_{hi}\right\}\right]^{-1}$, the diagonal matrix of the inverse of estimates of the theoretical variance of each cluster mean with variance components estimated from all the data. Such estimates of variance components from the data can be constrained to be positive or allowed to be negative.

If ρ is small and cluster sizes are small, $\boldsymbol{\Sigma}_2 \approx \left(\sigma_y^2\right)\left\{_d 1/n_{hi}\right\}$, so that weighting by cluster sizes is an appropriate choice. Note that specifying the diagonal elements of $\boldsymbol{W}$ as the inverse of cluster sizes following the inverse-variance paradigm given above is a common analytical error. If cluster sizes are not highly variable, $\boldsymbol{\Sigma}_2 \approx \sigma^2\boldsymbol{W}^{-1}$ directly, so an analysis of unweighted means is performed.

3.5 Description of Simulations

3.5.1 Tests Chosen

All simulations were performed with SAS version 9.2 PROC MIXED (one-stage analyses) or PROC GLM (two stage analyses). In line with the ideas presented in sections 3.3 and 3.4, this simulation study evaluated the following tests:

1. A one stage analysis of individual level data with inflation factor and denominator degrees of freedom calculated by the method of Kenward and Roger [13] and estimates of variance

components constrained to be positive. The Kenward and Roger [13] inflation factor and denominator degrees of freedom were achieved by using the DDFM=KR option on the MODEL statement in SAS PROC MIXED. By default, PROC MIXED constrains variance components to be positive.

2. A one stage analysis of individual level data with inflation factor and denominator degrees of freedom calculated by the method of Kenward and Roger [13] and estimates of variance components allowed to negative. Computation of negative estimates of variance components is allowed by use of the PARMS statement with NOBOUND option in PROC MIXED.

3. A one stage analysis of individual level data with denominator degrees of freedom equaling $m - g$ and estimates of variance components constrained to be positive. These degrees of freedom were obtained by use of the DDFM=BW option on the MODEL statement in SAS PROC MIXED with a RANDOM statement to specify random clusters. Note that the DDFM=BW option would not compute the correct degrees of freedom if used with a REPEATED statement.

4. A one stage analysis of individual level data with denominator degrees of freedom equaling $m - g$ and estimates of variance components allowed to be negative. As in test 2, computation of negative estimates of variance components is allowed by use of the PARMS statement with NOBOUND option in PROC MIXED.

5. A two stage analysis of cluster means with means unweighted, i.e., $\boldsymbol{W} = \boldsymbol{I}_m$.

6. A two stage analysis of cluster means with means weighted by cluster size, i.e., $\boldsymbol{W} = \left\{_d\, n_{hi}\right\}$.

7. A two stage analysis of cluster means with means weighted by the inverse of cluster size, i.e., $\boldsymbol{W} = \left\{_d\, 1/n_{hi}\right\}$.

8. A two stage analysis of cluster means with means weighted by the inverse of the estimated variance of each sample mean, i.e., $\boldsymbol{W} = \left\{_d [n_{hi}\,(n_{hi}-1)] \,/\, [\boldsymbol{y}'_{hi}\boldsymbol{y}_{hi} - n_{hi}\bar{y}_{1,hi}]\right\}$. Such weights were computed with PROC MEANS in SAS.

9. A two stage analysis of cluster means with means weighted by the inverse of the theoretical variance of a cluster mean, i.e. $\boldsymbol{W} = \left\{{}_d\left(\widehat{\sigma}_c^2 + \widehat{\sigma}_e^2/n_{hi}\right)^{-1}\right\}$. Variance components were estimated from the individual level data and were constrained to be positive. Estimates of variance components were obtained using PROC MIXED as in tests 1 and 3.

10. A two stage analysis of cluster means with means weighted by the inverse of the theoretical variance of a cluster mean, i.e. $\boldsymbol{W} = \left\{{}_d\left(\widehat{\sigma}_c^2 + \widehat{\sigma}_e^2/n_{hi}\right)^{-1}\right\}$. Variance components were estimated from the individual level data and were allowed to be negative. Estimates of variance components were obtained using PROC MIXED as in tests 2 and 4.

All computations assumed target $\alpha = 0.05$. A 95% confidence interval for type I error assuming true $\alpha = .05$ is $.05 \pm 1.96\sqrt{(.05)(.95)/N_c}$ where N_c is the number of simulated replications. For each of tests 1-10, 10,000 replications were simulated for each case of imbalance considered so that 95% confidence bounds around the type I error rate for each case are $\pm .0042$. Though type I error for tests 5 - 7 can be calculated exactly (without simulation) using theory from Chapter 2, type I errors for these tests were simulated so that they would be directly comparable to simulated type I errors for the other tests.

3.5.2 Description of Parameters of Imbalance

As in Chapter 2, we study a two group comparison, since most randomized trials involve two treatment arms [35]. We characterize cases of imbalance by six parameters: the number of clusters per treatment group, denoted as m_1 and m_2, the average cluster sizes of each treatment group, denoted as $\overline{n}_1$ and $\overline{n}_2$, the ratio of maximum to minimum cluster sizes for each treatment group, denoted as r_1 and r_2, and the within cluster correlation, ρ, which is assumed to be the same in both treatment groups.

3.5.3 Values of Parameters of Imbalance

Simulations in this paper focus on cases of imbalance common to group randomized trials. In such trials, due to randomization of clusters to treatment groups, the number of clusters per treatment group is always designed to be equal, though in some cases, treatment groups may vary by 1 or 2 clusters if entire clusters drop out of the study. To reflect this, these simulations

studied $m_1, m_2 \in \{(2,\,4),\ (3,\,4),\ (4,\,4),\ (6,\,8),\ (7,\,8),\ (8,\,8),\ (14,\,16),\ (15,\,16),\ (16{,}16)\}$. Cases considered for average number of observations per cluster per treatment group and ratio of maximum to minimum number of observations per cluster per treatment group were $\overline{n}_1$, $\overline{n}_2 \in \{8,\,16,\,32,\,64,\,128\}$ and $r_1,\ r_2 \in \{1,\,2,\,4,\,8\}$. Finally, this simulation study considered $\rho \in \{0.001,\,0.01,\,0.1\}$.

A full factorial combination of these parameters yields 18,000 cases of imbalance for each of 10 tests. Many of these lead to the same design with respect to type I error computation. For example, a case with $m_1, m_2, \overline{n}_1,\ \overline{n}_2, r_1, r_2 = \{4,\,4,\,8,\,16,\,1,\,1\}$ will have the same type I error as a case with $\overline{n}_1$ and $\overline{n}_2$ reversed. Unique cases can be characterized as those which have $(m_1 < m_2)$, or $(m_1 = m_2$ and $\overline{n}_1 < \overline{n}_2)$, or $(m_1 = m_2,\ \overline{n}_1 = \overline{n}_2$, and $r_1 \leq r_2)$. Of the 18,000 cases, 9,090 are unique. Type I error was computed only for unique cases.

3.5.4 Generation of Cluster Sizes

As in Chapter 2, cluster sizes were computed so they had distribution:

$$n_{hi} \sim \mathcal{N}\left\{\overline{n}_h,\ \left[2_h r_h / \left(r_h + 1\right)\right]^2\right\}.$$

Chapter 2 gives further information about how this distribution was chosen. Define the constant $c_2 = \text{Prob}\left\{(Z < -2)\right\}$, where $Z \sim \mathcal{N}(0, 1)$. Cluster sizes $\{n_{hi}\}$ were computed as:

$$n_{hi} = \mathcal{Z}^{-1}\left[c_2 + (i-1)\left(1 - c_2\right)/m_h\right]\sigma + \overline{n}_h$$

for $h = 1, 2$ and $i = 1, \ldots, m_h$, where $\mathcal{Z}^{-1}(p)$ denotes a function with returns the p-th quantile from a standard normal distribution. Cluster sizes were rounded to the nearest whole number so that actual ratios for maximum to minimum cluster size were not achieved.

3.5.5 Convergence in Simulations

Computations for tests 5-8 were non-iterative. When variance components were constrained to be positive, as in tests 1, 3, and 9, PROC MIXED always converged, so that 10,000 replications were realized for all cases of imbalance for these tests.

When variance components were allowed to be negative, the simulations often did not converge for cases where negative estimates would have occurred (i.e. when the cluster variance component was estimated to be zero in the constrained analysis). We originally ran simulations

with default (MIVQUE) starting values, but later found that using ordinary least squares start values (specified with the OLS option on the PARMS statement) lead to better convergence.

Tables 3.1 and 3.2 show descriptive statistics for the number of converged replications and number of replications with a negative cluster variance component, respectively, for tests 2, 4, and 10 by number of clusters and ρ. Low rates of convergence often occurred for small number of clusters, $m_1, m_2 \in \{(2,4),(3,4)\}$. Of these, the largest rates of non-convergence occurred when both $\overline{n}_1 > \overline{n}_2$ and $r_1 > r_2$ (results not shown), that is, when the treatment group with the smaller number of clusters had the larger and more variable cluster sizes.

3.6 Results of Simulation Study

3.6.1 Overview

With ten tests and 9,090 cases of imbalance per test considered, type I error for each case of imbalance for each test cannot be presented separately. We confine discussion to type I error for main effects of parameters of imbalance or combinations of them.

A 95% confidence interval for the type I error of each case of imbalance was computed as:

$$\widehat{\alpha} \pm 1.96\sqrt{(\widehat{\alpha})(1-\widehat{\alpha})/N_c},$$

where $\widehat{\alpha}$ is the simulation type I error and N_c is the number of simulation replications observed for that case. Type I error was considered approximately unbiased if this interval overlapped with the interval (.04, .06), a 20% difference from nominal α, that is, if:

$$\widehat{\alpha} \in \left(.04 - 1.96\sqrt{(\widehat{\alpha})(1-\widehat{\alpha})/N_c}, .06 + 1.96\sqrt{(\widehat{\alpha})(1-\widehat{\alpha})/N_c}\right).$$

3.6.2 Summary Over All Cases

Table 3.3 shows descriptive statistics for type I error over all cases of imbalance. Over all parameters, test 9 controlled type I error well in 8,802 (97%) of all cases. Further tables will discuss scenarios under which test 9 showed biased type I error. Over all cases, test 9 controlled test size for 808, 993, and 1,081 more cases than the next best tests, tests 1, 2, and 5, respectively. Test 3 alone was conservative for virtually all cases, though also extremely so in many cases. All tests were unbiased in some circumstances; subsequent tables discuss this.

3.6.3 Pairwise Comparison of Tests Over All Cases

Table 3.4 shows the number and percent of cases in which pair-wise combinations of tests have one test biased and the other unbiased. Table 3.5 shows the positive differences of pair-wise combinations. No one test was always superior over another. That is, for no pair wise combinations of tests is one test always unbiased when the other is biased; however, test 9 was unbiased when all other tests were biased for more cases. In particular, Test 9 was unbiased and Test 1 biased for 1,033 (11%) of cases; conversely, Test 1 was unbiased and Test 9 biased for 225 (2%) of cases.

3.6.4 Type I Error by Within Cluster Correlation

Table 3.6 gives type I error by value of within cluster correlation, ρ. Tests 6 and 9 were always unbiased when $\rho = .001$. This finding for test 6 agrees with the results from Chapter 2. No other test controlled type I error well when $\rho = .001$. Tests 9 and 1 had unbiased type I error for roughly equivalent number of cases, 2,937 (97%) and 2,925 (97%), respectively when $\rho = .01$. Test 1 controlled type I error the best when $\rho = .1$, with 76 more cases that test 9. All tests with the exception of tests 6, 7, and 8 controlled type I error in more than 90% of cases when $\rho = .1$. In general, as ρ increased, tests 6, 8, and 9 became more biased and tests 1, 2, 3, 4, 5, 7, and 10 became less biased.

3.6.5 Type I Error by Number of Clusters

Table 3.7 gives type I error by value of $m_1 + m_2$. By design, $m_2 - m_1 \in (0, 1, 2)$ and $m_2 \in (4, 8, 16)$. With 4 or fewer clusters, test 9 controlled type I error in 91% of cases; no other test controlled type I error well for a small number of clusters. Tests 1, 2, 9, and 10 controlled type I error for more than 90% of cases when $m_2 = 8$. These same tests as well as tests 3, 4, and 5 had unbiased type I error for more than 90% of cases when $m_2 = 16$.

3.6.6 Type I Error For Selected Scenarios of Balance

Tables 3.8 and 3.9 give type I error for the following scenarios of balanced number of cluster and/or number of observations per cluster, common to many group randomized trials:

1. Average cluster sizes per treatment group are equal and all clusters sizes within a treat-

ment group are equal (i.e., balanced data where a hypothesis test with nominal α is available).

2. Numbers of clusters per treatment group are equal, average cluster sizes per treatment group vary by no more than a factor or two, and ratios of maximum to minimum cluster size per treatment group are less than two.

3. Numbers of clusters per treatment group are equal, average cluster sizes per treatment group vary by no more than a factor or two, and at least one ratio of maximum to minimum cluster size per treatment group is larger than two.

4. Numbers of clusters per treatment group are unequal, average cluster sizes per treatment group vary by no more than a factor or two, and ratios of maximum to minimum cluster size per treatment group are less than two.

5. Numbers of clusters per treatment group are unequal, average cluster sizes per treatment group vary by no more than a factor or two, and at least one ratio of maximum to minimum cluster size per treatment group is larger than two.

6. Numbers of clusters and average cluster sizes per treatment group are equal.

7. Numbers of clusters per treatment group are equal.

8. Average cluster sizes per treatment group are equal.

Tests 1, 3, and 8 were often biased even for completely balanced data, as expected. Test 9 was unbiased for all types of balance mentioned. Test 5, though biased in a few cases, was also largely unbiased ($> 94\%$ of cases) for all these types of balance. Test 6 controlled type I error well when the average cluster sizes per treatment group were close to equal. Test 4 was unbiased when the number of clusters were equal and the average cluster sizes were close to balanced. Tests 1 and 2 were unbiased when the number of clusters were equal and the average cluster sizes were unbalanced. Both tests performed worse for balanced average cluster sizes. Tests 3, 7, 8 were often biased for all types of balance considered, except as mentioned previously for completely balanced data.

3.6.7 Decision Tables for Type I Error

Figures 3.1 - 3.10 provide pictorial representations of when type I error is unbiased for each test for combinations of ρ, m_1, m_2, $\overline{n}_1$, and $\overline{n}_2$. As in Chapter 2, ratio of maximum to minimum cluster size was the least important factor in contributing to bias in type I error; for brevity, results are summarized over r_1 and r_2. The first three rows are shaded as follows: black when all cases of $r_1 \times r_2$ have unbiased type I error, gray when some cases (number indicated in the cell) of $r_1 \times r_2$ have unbiased type I error, and white when no cases of $r_1 \times r_2$ have unbiased type I error. The second three rows show median type I error times 1000 ($.05 = 50$) for each combination of ρ, m_1, m_2, $\overline{n}_1$, and $\overline{n}_2$.

3.7 Conclusions

In this paper, type I error was simulated for several tests of fixed effects in the analysis of group randomized trials data. Except in very extreme cases and small number of clusters, the analysis of cluster means with weights $\boldsymbol{W} = \left\{{}_d\left(\widehat{\sigma}_c^2 + \widehat{\sigma}_e^2/n_{hi}\right)^{-1}\right\}$, where estimates $\widehat{\sigma}_c^2$ and $\widehat{\sigma}_e^2$ are constrained to be positive, controls type I error and should be recommend as the analysis of choice for unbalanced clustered data. Computation of power for this analysis with unbalanced clustered data is discussed in Chapter 4.

A one-stage model with inflation factor and degrees of freedom computed by the method of Kenward and Roger [13] (with variance components constrained to be either positive or negative) controlled type I error for the second and third most number of cases. These tests are not recommended in general, however, since they are most biased for cases that were nearly balanced. Convergence also often was not achieved for the test with negative variance components. Further, no methods exist to compute power for these tests.

The two stage analysis of means unweighted and weighted by cluster size also controlled type I error in a subset of cases: balanced number of clusters or balanced cluster size or large within cluster correlation for the former and balanced cluster size or small within cluster correlation for the latter.

When all of number of clusters, average cluster size, and ratio of maximum to minimum cluster size are close to balanced or when the number of clusters is large, several of the methods controlled type I error well. This situation is common to data collected in most primary analyses

of large group randomized trials, so that this research confirms that for most group randomized trials, an unbiased test is available using almost all methods. In particular, the one-stage model with denominator degrees of freedom equaling $m - g$, most often used in group randomized trials, is included in the tests that performed well in these scenarios.

Future research will explore type I error of these methods for non-randomized studies, which often have more unbalanced numbers of clusters, i.e., $|m_1 - m_2| > 2$. Future research will also consider both randomized and non-randomized designs with treatment groups that have unequal within cluster correlation.

Table 3.1: Number of convergent scenarios for tests 2, 4, and 10, by ρ and $m_1 \times m_2$

ρ	m_1	m_2	N[1]	Min	Q1	Med	Q3	Max
.001	2	4	400	4,263	6,377	8,533	9,222	10,000
	3	4	400	6,378	8,048	8,709	9,353	10,000
	4	4	210	8,244	8,658	9,214	9,763	10,000
	6	8	400	8,611	9,499	9,686	9,932	10,000
	7	8	400	8,903	9,516	9,735	9,943	10,000
	8	8	210	9,314	9,656	9,868	9,981	10,000
	14	16	400	9,749	9,922	9,981	9,999	10,000
	15	16	400	9,774	9,916	9,984	9,999	10,000
	16	16	210	9,831	9,953	9,992	10,000	10,000
.01	2	4	400	5,036	7,566	9,126	9,655	10,000
	3	4	400	7,037	8,612	9,244	9,745	10,000
	4	4	210	8,681	9,212	9,517	9,892	10,000
	6	8	400	9,029	9,759	9,918	9,991	10,000
	7	8	400	9,274	9,811	9,931	9,993	10,000
	8	8	210	9,594	9,891	9,955	9,997	10,000
	14	16	400	9,861	9,991	9,999	10,000	10,000
	15	16	400	9,890	9,991	9,999	10,000	10,000
	16	16	210	9,951	9,996	10,000	10,000	10,000
.1	2	4	400	7,812	9,415	9,865	9,978	10,000
	3	4	400	8,843	9,754	9,918	9,992	10,000
	4	4	210	9,594	9,893	9,966	9,999	10,000
	6	8	400	9,868	9,995	10,000	10,000	10,000
	7	8	400	9,913	9,997	10,000	10,000	10,000
	8	8	210	9,964	9,999	10,000	10,000	10,000
	14	16	400	9997	10,000	10,000	10,000	10,000
	15	16	400	9999	10,000	10,000	10,000	10,000
	16	16	210	9999	10,000	10,000	10,000	10,000

[1] N = Number of simulation scenarios

Table 3.2: Number of scenarios with a negative variance component estimate, by ρ and $m_1 \times m_2$

ρ	m_1	m_2	N[1]	Min	Q1	Med	Q3	Max
.001	2	4	400	5,327	5,763	5,904	6,056	6,515
	3	4	400	5,118	5,680	5,825	5,949	6,332
	4	4	210	4,940	5,538	5,692	5,830	6,062
	6	8	400	4,289	4,936	5,283	5,457	5,766
	7	8	400	4,240	4,904	5,246	5,419	5,749
	8	8	210	4,175	4,834	5,227	5,382	5,619
	14	16	400	3,592	4,280	4,786	5,103	5,369
	15	16	400	3,543	4,223	4,774	5,071	5,365
	16	16	210	3,517	4,184	4,765	5,124	5,360
.01	2	4	400	2,228	3,595	4,679	5,216	5,667
	3	4	400	1,811	3,226	4,077	4,838	5,524
	4	4	210	1,445	2,825	3,809	4,655	5,301
	6	8	400	509	1,507	2,496	3,707	4,864
	7	8	400	422	1,326	2,465	3,637	4,872
	8	8	210	370	1,204	2,415	3,511	4,721
	14	16	400	34	350	1,233	2,537	4,307
	15	16	400	34	318	1,213	2,454	4,304
	16	16	210	27	300	1,171	2,502	4,279
.1	2	4	400	76	384	914	1,881	3,369
	3	4	400	23	238	578	1,169	2,864
	4	4	210	7	165	353	854	2,428
	6	8	400	0	3	30	137	1,178
	7	8	400	0	3	16	101	1,144
	8	8	210	0	2	10	92	915
	14	16	400	0	0	0	2	279
	15	16	400	0	0	0	1	239
	16	16	210	0	0	0	1	229

[1] N = Number of simulation scenarios

Table 3.3: Type I error over all cases

Test	N[1]	Min	Q1	Med	Q3	Max	R[2]	N (%)[3]
9	**9,090**	**.019**	**.048**	**.050**	**.052**	**.087**	**.068**	**8,802(97%)**
1	9,090	.027	.041	.047	.051	.092	.065	7,994 (88%)
2	9,090	<.001	.045	.049	.052	.100	.100	7,809 (86%)
5	9,090	.017	.047	.050	.055	.215	.198	7,721 (85%)
10	9,090	.005	.046	.050	.054	.171	.166	7,275 (80%)
6	9,090	<.001	.042	.049	.052	.131	.131	6,964 (77%)
4	9,090	<.001	.038	.047	.051	.160	.159	6,771 (74%)
3	9,090	.003	.026	.038	.047	.066	.063	4,934 (54%)
8	9,090	.008	.053	.061	.079	.236	.228	4,688 (52%)
7	9,090	<.001	.017	.049	.083	.375	.375	2,335 (26%)

[1] N = Number of simulation scenarios
[2] Range = Max − Min
[3] Number (%) of cases with $\alpha \in (.04 - 95\%\text{CI STE}, .06 + 95\%\text{CI STE})$

Table 3.4: Number (%) of total (N=9,090) cases where test i (down) is biased and test j (across) is unbiased.

(i,j)	1	2	3	4	5	6	7	8	9	10
1		481 (5%)	93 (1%)	556 (6%)	822 (9%)	1,029 (11%)	271 (3%)	671 (7%)	1,033 (11%)	490 (5%)
2	666 (7%)		221 (2%)	620 (7%)	936 (10%)	1,044 (11%)	253 (3%)	705 (8%)	1,145 (13%)	598 (7%)
3	3,153 (35%)	3,096 (34%)		1,918 (21%)	2,929 (32%)	3,662 (40%)	968 (11%)	2,213 (24%)	3,981 (44%)	2,419 (27%)
4	1,779 (20%)	1,658 (18%)	81 (1%)		1,186 (13%)	1,815 (20%)	338 (4%)	1,123 (12%)	2,107 (23%)	919 (10%)
5	1,095 (12%)	1,024 (11%)	142 (2%)	236 (3%)		1,037 (11%)	190 (2%)	614 (7%)	1,221 (13%)	668 (7%)
6	2,059 (23%)	1,889 (21%)	1,632 (18%)	1,622 (18%)	1,794 (20%)		250 (3%)	635 (7%)	1,852 (20%)	1,890 (21%)
7	5,930 (65%)	5,727 (63%)	3,567 (39%)	4,774 (53%)	5,576 (61%)	4,879 (54%)		3,164 (35%)	6,501 (72%)	5,205 (57%)
8	3,977 (44%)	3,826 (42%)	2,459 (27%)	3,206 (35%)	3,647 (40%)	2,911 (32%)	811 (9%)		4,154 (46%)	3,550 (39%)
9	225 (2%)	152 (2%)	113 (1%)	76 (1%)	140 (2%)	14 (0%)	34 (0%)	40 (0%)		124 (1%)
10	1,209 (13%)	1,132 (12%)	78 (1%)	415 (5%)	1,114 (12%)	1,579 (17%)	265 (3%)	963 (11%)	1,651 (18%)	

Table 3.5: Difference in Number (i,j) and Number (j,i) from Table 3.4.

(i,j)	1	2	3	4	5	6	7	8	9	10
1									808	
2	185								993	
3	3,060	2,875		1,837	2,787	2,030			3,868	2,341
4	1,223	1,038			950	193			2,031	504
5	273	88							1,081	
6	1,030	845			757				1,838	311
7	5,659	5,474	2,599	4,436	5,386	4,629		2,353	6,467	4,940
8	3,306	3,121	246	2,083	3,033	2,276			4,114	2,587
9										2,587
10	719	534			446				1,527	

Only positive differences are shown.

Table 3.6: Type I error by ρ

	Test	N[1]	Min	Q1	Med	Q3	Max	R[2]	N (%)[3]	
$\rho = .001$	**6**	**3,030**	**.040**	**.048**	**.050**	**.051**	**.058**	**.018**	**3,030**	**(100%)**
	9	**3,030**	**.044**	**.049**	**.051**	**.053**	**.059**	**.015**	**3,030**	**(100%)**
	2	3,030	<.001	.041	.048	.050	.097	.097	2,474	(82%)
	5	3,030	.017	.045	.050	.059	.215	.198	2,317	(76%)
	1	3,030	.027	.036	.040	.044	.053	.026	2,132	(70%)
	10	3,030	.005	.045	.051	.061	.171	.166	2,045	(67%)
	4	3,030	<.001	.031	.042	.049	.160	.159	1,861	(61%)
	8	3,030	.046	.055	.064	.091	.236	.190	1,609	(53%)
	3	3,030	.003	.011	.027	.036	.045	.041	734	(24%)
	7	3,030	<.001	.015	.049	.090	.375	.375	694	(23%)
$\rho = .01$	**9**	**3,030**	**.027**	**.047**	**.050**	**.052**	**.078**	**.051**	**2,937**	**(97%)**
	1	**3,030**	**.030**	**.043**	**.047**	**.050**	**.071**	**.041**	**2,925**	**(97%)**
	2	3,030	<.001	.045	.049	.051	.094	.094	2,577	(85%)
	5	3,030	.018	.046	.050	.057	.167	.149	2,489	(82%)
	6	3,030	.018	.039	.048	.052	.087	.069	2,373	(78%)
	10	3,030	.008	.045	.050	.054	.167	.159	2,369	(78%)
	4	3,030	.005	.035	.045	.050	.155	.149	2,113	(70%)
	8	3,030	.031	.053	.061	.078	.235	.204	1,817	(60%)
	3	3,030	.005	.019	.035	.043	.056	.051	1,431	(47%)
	7	3,030	<.001	.016	.049	.086	.346	.346	741	(24%)
$\rho = .1$	**1**	**3,030**	**.036**	**.049**	**.051**	**.054**	**.092**	**.056**	**2,937**	**(97%)**
	5	**3,030**	**.027**	**.048**	**.050**	**.053**	**.086**	**.059**	**2,915**	**(96%)**
	10	**3,030**	**.018**	**.048**	**.050**	**.052**	**.119**	**.101**	**2,861**	**(94%)**
	9	**3,030**	**.019**	**.048**	**.050**	**.053**	**.087**	**.068**	**2,835**	**(94%)**
	4	**3,030**	**.016**	**.046**	**.049**	**.052**	**.113**	**.097**	**2,797**	**(92%)**
	3	**3,030**	**.016**	**.046**	**.049**	**.051**	**.066**	**.050**	**2,269**	**(91%)**
	2	**3,030**	**<.001**	**.048**	**.050**	**.053**	**.100**	**.100**	**2,758**	**(91%)**
	6	3,030	<.001	.025	.045	.056	.131	.131	1,561	(52%)
	8	3,030	.008	.041	.058	.074	.222	.214	1,262	(42%)
	7	3,030	<.001	.019	.049	.076	.278	.278	900	(30%)

[1] N = Number of simulation scenarios

[2] Range = Max − Min

[3] Number (%) of cases with $\alpha \in (.04 - 95\%\text{CI STE}, .06 + 95\%\text{CI STE})$

Table 3.7: Type I error by $m_1 + m_2$

.

	Test	N[1]	Min	Q1	Med	Q3	Max	R[2]	N (%)[3]	
$m_1 + m_2 = 6, 7, 8$	**9**	**3,030**	**.019**	**.048**	**.051**	**.054**	**.087**	**.068**	**2,744**	**(91%)**
	6	3,030	.004	.046	.050	.055	.131	.127	2,318	(77%)
	1	3,030	.027	.037	.045	.053	.092	.065	2,229	(74%)
	5	3,030	.017	.043	.050	.059	.215	.198	2,093	(69%)
	2	3,030	.000	.030	.044	.052	.100	.100	1,868	(62%)
	8	3,030	.008	.053	.060	.073	.179	.171	1,694	(56%)
	10	3,030	.005	.038	.051	.068	.171	.166	1,484	(49%)
	4	3,030	.001	.018	.035	.050	.160	.159	1,286	(42%)
	3	3,030	.003	.009	.015	.037	.066	.063	773	(26%)
	7	3,030	.000	.030	.064	.112	.375	.374	664	(22%)
$m_1 + m_2 = 14, 15, 16$	**9**	**3,030**	**.035**	**.048**	**.050**	**.052**	**.059**	**.024**	**3,028**	**(100%)**
	2	**3,030**	**.025**	**.046**	**.049**	**.051**	**.058**	**.033**	**2,911**	**(96%)**
	10	**3,030**	**.032**	**.046**	**.050**	**.054**	**.084**	**.052**	**2,761**	**(91%)**
	1	**3,030**	**.030**	**.041**	**.047**	**.050**	**.058**	**.028**	**2,740**	**(90%)**
	5	3,030	.032	.047	.050	.056	.106	.075	2,683	(89%)
	4	3,030	.026	.039	.047	.051	.081	.054	2,455	(81%)
	6	3,030	.002	.040	.048	.051	.077	.075	2,366	(78%)
	8	3,030	.009	.053	.061	.081	.214	.205	1,577	(52%)
	3	3,030	.021	.028	.035	.047	.059	.038	1,412	(47%)
	7	3,030	.000	.016	.048	.076	.234	.234	842	(28%)
$m_1 + m_2 = 30, 31, 32$	**2**	**3,030**	**.043**	**.049**	**.050**	**.052**	**.057**	**.014**	**3,030**	**(100%)**
	4	**3,030**	**.038**	**.045**	**.049**	**.051**	**.060**	**.023**	**3,030**	**(100%)**
	9	**3,030**	**.042**	**.048**	**.050**	**.052**	**.058**	**.016**	**3,030**	**(100%)**
	10	**3,030**	**.040**	**.047**	**.050**	**.052**	**.062**	**.022**	**3,030**	**(100%)**
	1	**3,030**	**.035**	**.045**	**.048**	**.051**	**.057**	**.022**	**3,025**	**(100%)**
	5	**3,030**	**.040**	**.048**	**.051**	**.054**	**.077**	**.037**	**2,945**	**(97%)**
	3	**3,030**	**.031**	**.039**	**.044**	**.049**	**.056**	**.026**	**2,749**	**(91%)**
	6	3,030	.001	.037	.047	.051	.068	.067	2,280	(75%)
	8	3,030	.010	.053	.062	.091	.236	.226	1,417	(47%)
	7	3,030	.000	.009	.039	.065	.190	.190	829	(27%)

[1] N = Number of simulation scenarios
[2] Range = Max − Min
[3] Number (%) of cases with $\alpha \in (.04 - 95\%\text{CI STE}, .06 + 95\%\text{CI STE})$

Table 3.8: Type I error for selected scenarios of balance

	Test	N[1]	Min	Q1	Med	Q3	Max	R[2]	N (%)[3]
$\bar{n}_1 = \bar{n}_2$,	**2,4,5,6,**	**135**	**.045**	**.049**	**.050**	**.052**	**.055**	**.010**	**135 (100%)**
$r_1 = r_2 = 1$	**7,9,10**								
(Balanced data)	1	135	.030	.037	.045	.050	.055	.026	109 (81%)
	8	135	.049	.053	.057	.067	.100	.051	98 (73%)
	3	135	.004	.026	.037	.048	.055	.052	72 (53%)
$m_1 = m_2$	**5**	**279**	**.045**	**.049**	**.051**	**.052**	**.058**	**.013**	**279 (100%)**
$\bar{n}_2/\bar{n}_1 \in (.5, 1, 2)$	**6**	**279**	**.040**	**.048**	**.050**	**.052**	**.058**	**.018**	**279 (100%)**
$r_1, r_2 \in (1, 2)$	**9**	**279**	**.045**	**.049**	**.050**	**.052**	**.056**	**.011**	**279 (100%)**
	2	**279**	**.027**	**.048**	**.050**	**.052**	**.061**	**.034**	**264 (95%)**
	7	**279**	**.030**	**.041**	**.049**	**.054**	**.061**	**.032**	**260 (93%)**
	4	**279**	**.033**	**.047**	**.050**	**.052**	**.075**	**.042**	**259 (93%)**
	10	**279**	**.036**	**.048**	**.051**	**.053**	**.079**	**.043**	**256 (92%)**
	1	279	.029	.039	.046	.050	.056	.027	235 (84%)
	8	279	.043	.053	.057	.067	.109	.065	197 (71%)
	3	279	.011	.027	.039	.048	.056	.045	155 (56%)
$m_1 = m_2$	**9**	**747**	**.044**	**.050**	**.051**	**.053**	**.065**	**.021**	**746 (100%)**
$\bar{n}_2/\bar{n}_1 \in (.5, 1, 2)$	**5**	**747**	**.036**	**.048**	**.050**	**.051**	**.056**	**.020**	**744 (100%)**
$r_1 \in (4, 8)$ or	**6**	**747**	**.042**	**.050**	**.052**	**.055**	**.089**	**.046**	**712 (95%)**
$r_2 \in (4, 8)$	**4**	**747**	**.029**	**.049**	**.051**	**.054**	**.124**	**.096**	**675 (90%)**
	2	747	.025	.042	.048	.051	.062	.036	655 (88%)
	1	747	.028	.039	.046	.050	.056	.028	629 (84%)
	10	747	.044	.051	.055	.069	.131	.087	529 (71%)
	3	747	.010	.029	.040	.049	.060	.050	446 (60%)
	8	747	.046	.056	.062	.073	.228	.183	446 (60%)
	7	747	.039	.063	.083	.111	.250	.211	200 (27%)
$m_1 \neq m_2$	**9**	**936**	**.039**	**.048**	**.050**	**.052**	**.063**	**.024**	**936 (100%)**
$\bar{n}_2/\bar{n}_1 \in (.5, 1, 2)$	**6**	**936**	**.032**	**.047**	**.050**	**.052**	**.072**	**.040**	**906 (97%)**
$r_1, r_2 \in (1, 2)$	**5**	**936**	**.034**	**.048**	**.050**	**.053**	**.081**	**.047**	**881 (94%)**
	10	936	.019	.047	.050	.052	.109	.091	835 (89%)
	4	936	.013	.045	.049	.052	.105	.092	813 (87%)
	2	936	<.001	.047	.050	.052	.089	.089	809 (86%)
	1	936	.027	.038	.045	.050	.061	.034	762 (81%)
	8	936	.033	.053	.057	.067	.108	.075	676 (72%)
	7	936	.018	.036	.049	.056	.126	.108	609 (65%)
	3	936	.004	.025	.037	.048	.057	.053	488 (52%)

[1] N = Number of simulation scenarios
[2] Range = Max − Min
[3] Number (%) of cases with $\alpha \in (.04 - 95\%\text{CI STE}, .06 + 95\%\text{CI STE})$

Table 3.9: Type I error for selected scenarios of balance (cont.)

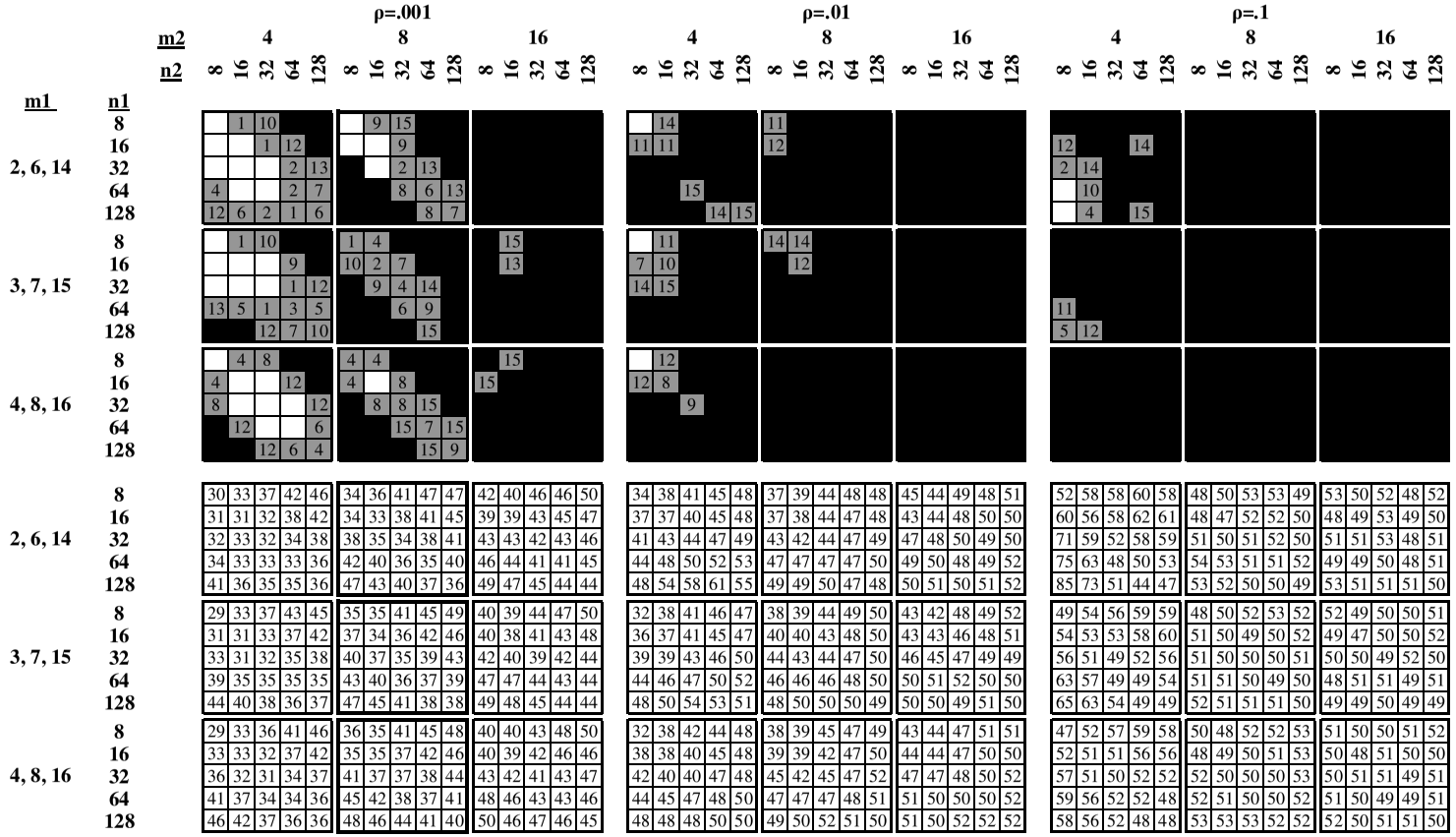
	Test	N[1]	Min	Q1	Med	Q3	Max	R[2]	N (%)[3]	
$m_1 \neq m_2$	**9**	**2,808**	**.043**	**.049**	**.051**	**.053**	**.083**	**.040**	**2,741**	**(98%)**
$\bar{n}_2/\bar{n}_1 \in (.5, 1, 2)$	**5**	**2,808**	**.024**	**.046**	**.049**	**.053**	**.082**	**.058**	**2,647**	**(94%)**
$r_1 \in (4, 8)$ or	**6**	**2,808**	**.036**	**.049**	**.052**	**.055**	**.131**	**.095**	**2,588**	**(92%)**
$r_2 \in (4, 8)$	4	2,808	.011	.046	.050	.053	.160	.149	2,377	(85%)
	1	2,808	.027	.038	.045	.050	.074	.047	2,289	(82%)
	10	2,808	.012	.050	.052	.058	.171	.160	2,241	(80%)
	2	2,808	<.001	.038	.047	.050	.097	.097	2,133	(76%)
	8	2,808	.039	.055	.061	.073	.236	.197	1,693	(60%)
	3	2,808	.003	.025	.038	.049	.065	.062	1,531	(55%)
	7	2,808	.015	.062	.087	.121	.375	.359	698	(25%)
$m_1 = m_2$	**5**	**450**	**.036**	**.048**	**.050**	**.051**	**.055**	**.019**	**449**	**(100%)**
$\bar{n}_2 = \bar{n}_1$	**9**	**450**	**.045**	**.050**	**.051**	**.053**	**.065**	**.020**	**449**	**(100%)**
	6	**450**	**.045**	**.051**	**.053**	**.056**	**.089**	**.043**	**421**	**(94%)**
	4	450	.045	.051	.054	.058	.124	.079	394	(88%)
	1	450	.028	.038	.046	.050	.055	.027	369	(82%)
	2	450	.025	.040	.048	.051	.055	.030	366	(81%)
	10	450	.045	.051	.054	.065	.131	.086	337	(75%)
	3	450	.010	.029	.040	.050	.060	.049	264	(59%)
	8	450	.048	.056	.062	.072	.228	.180	260	(58%)
	7	450	.045	.058	.079	.112	.250	.205	150	(33%)
$m_1 = m_2$	**9**	**1,890**	**.037**	**.049**	**.051**	**.052**	**.065**	**.028**	**1,889**	**(100%)**
	5	**1,890**	**.036**	**.049**	**.051**	**.054**	**.080**	**.044**	**1,847**	**(98%)**
	2	**1,890**	**.025**	**.047**	**.050**	**.052**	**.062**	**.037**	**1,781**	**(94%)**
	1	**1,890**	**.028**	**.042**	**.048**	**.051**	**.063**	**.035**	**1,712**	**(91%)**
	4	1,890	.014	.041	.048	.051	.124	.110	1,520	(80%)
	6	1,890	.002	.042	.049	.052	.089	.087	1,488	(79%)
	10	1,890	.025	.049	.052	.060	.131	.107	1,472	(78%)
	3	1,890	.010	.029	.039	.048	.060	.050	1,096	(58%)
	8	1,890	.012	.053	.060	.079	.228	.217	995	(53%)
	7	1,890	.000	.017	.047	.073	.250	.250	580	(31%)
$\bar{n}_2 = \bar{n}_1$	**5**	**1,890**	**.035**	**.047**	**.049**	**.051**	**.057**	**.023**	**1,879**	**(99%)**
	9	**1,890**	**.045**	**.050**	**.051**	**.053**	**.074**	**.029**	**1,868**	**(99%)**
	6	**1,890**	**.045**	**.050**	**.053**	**.057**	**.110**	**.065**	**1,755**	**(93%)**
	4	1,890	.030	.050	.053	.056	.160	.130	1,689	(89%)
	10	1,890	.033	.051	.054	.061	.171	.138	1,500	(79%)
	1	1,890	.027	.037	.045	.050	.067	.040	1,499	(79%)
	2	1,890	.000	.036	.047	.050	.089	.089	1,397	(74%)
	8	1,890	.047	.056	.061	.072	.236	.189	1,138	(60%)
	3	1,890	.003	.026	.039	.049	.064	.061	1,042	(55%)
	7	1,890	.045	.059	.080	.117	.291	.245	632	(33%)

[1] N = Number of simulation scenarios

[2] Range = Max − Min

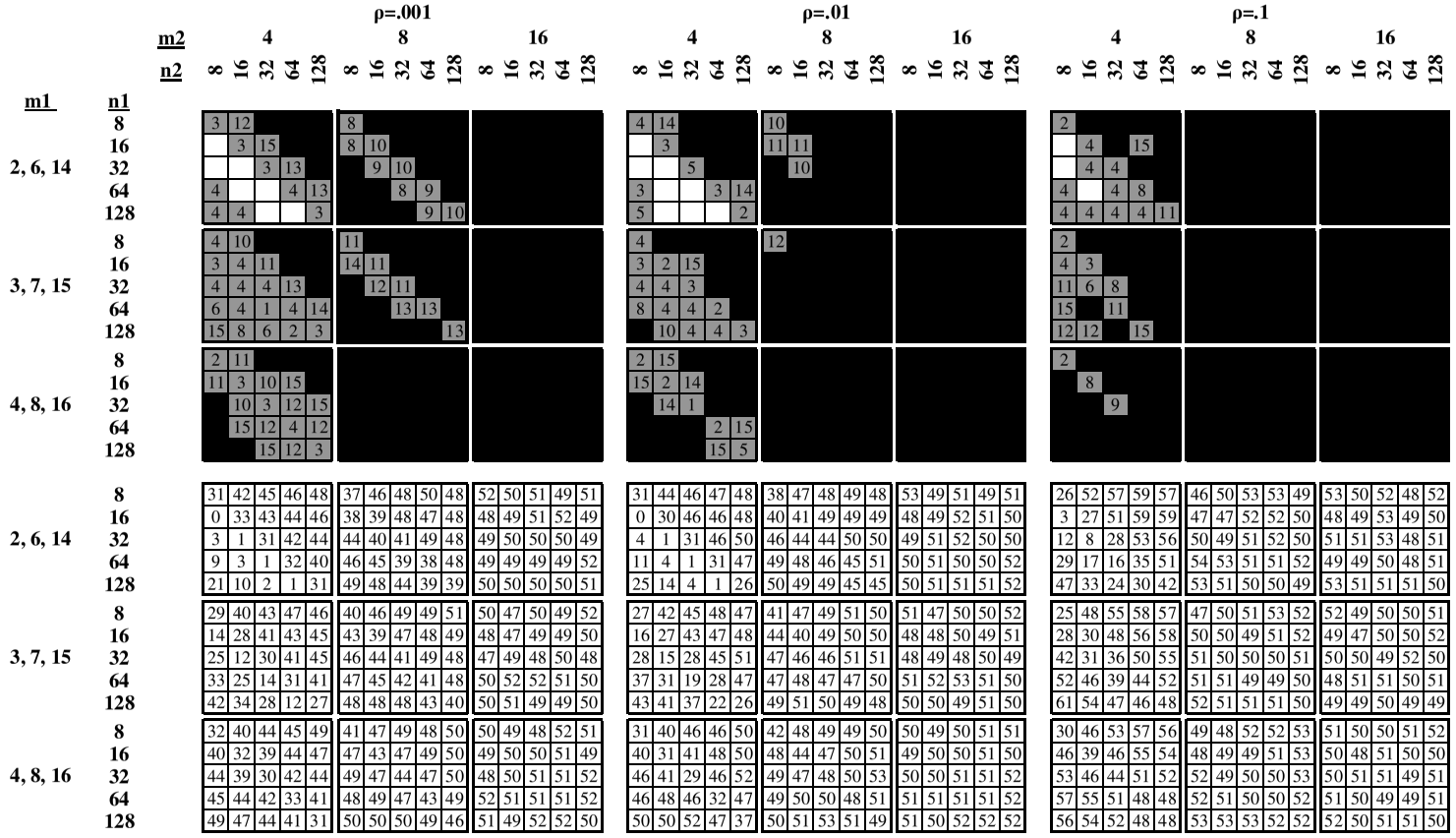
[3] Number (%) of cases with $\alpha \in (.04 - 95\%\text{CI STE}, .06 + 95\%\text{CI STE})$

Figure 3.1: Type I Error for Test 1 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



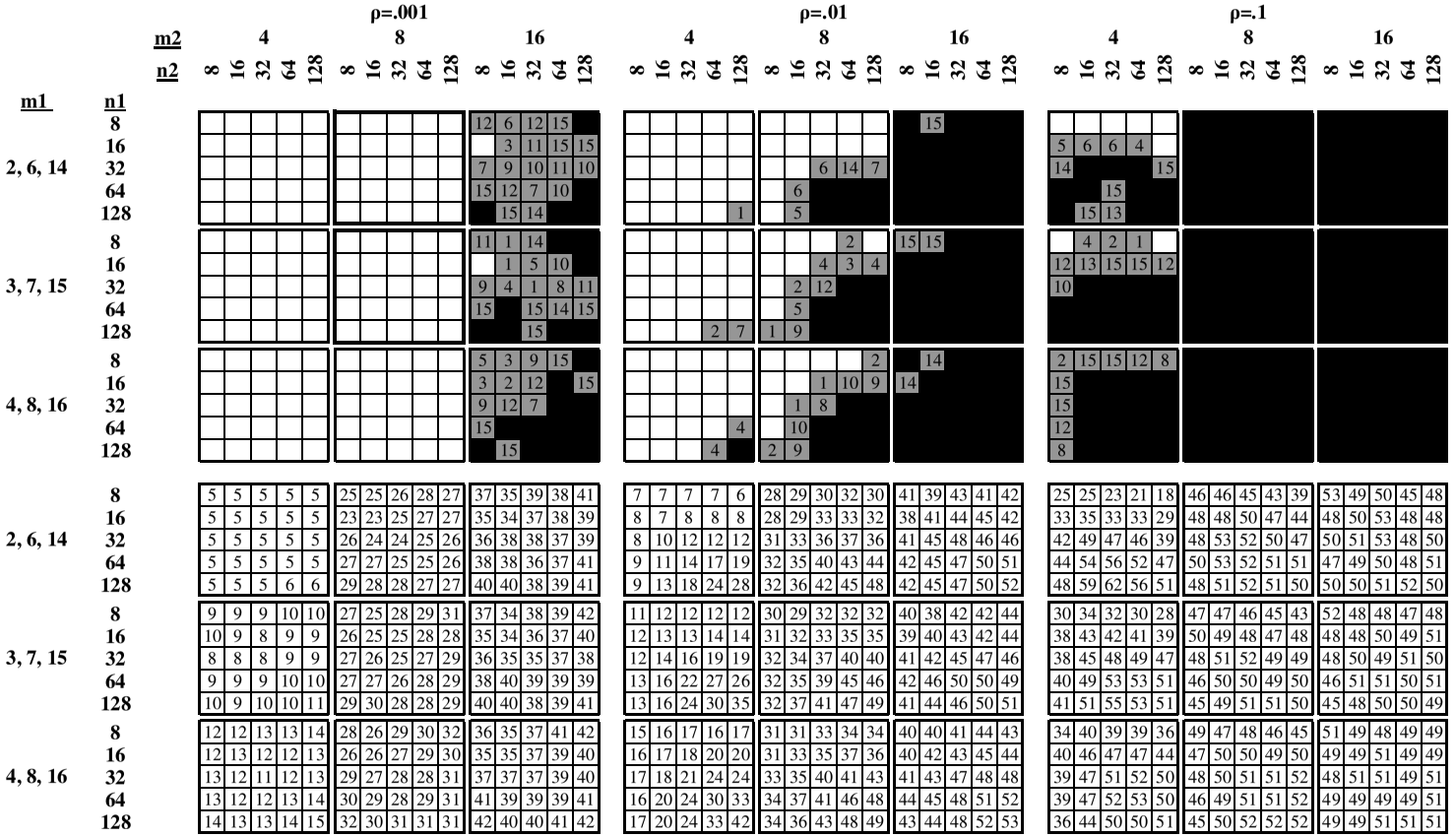
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.2: Type I Error for Test 2 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



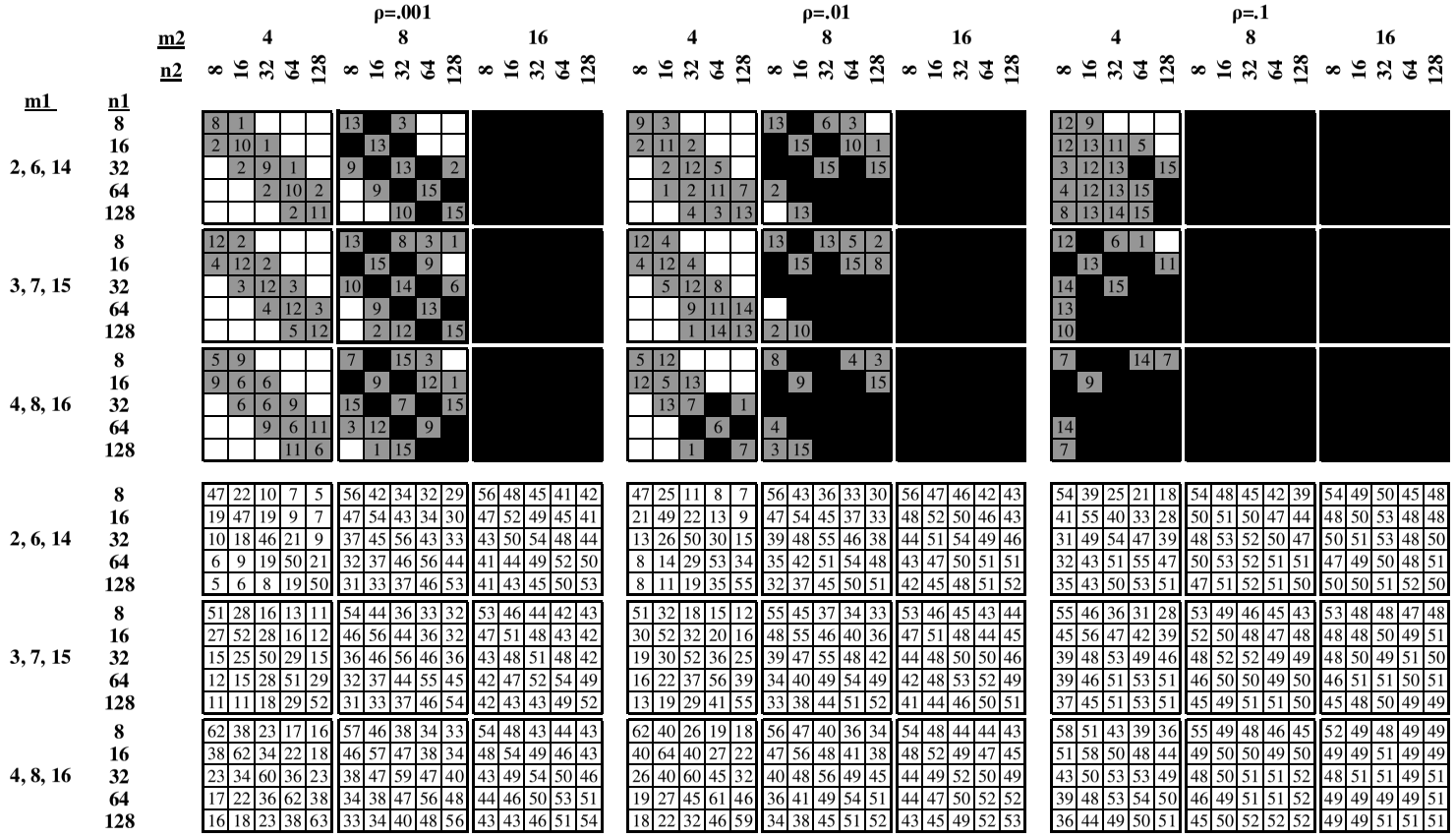
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.3: Type I Error for Test 3 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



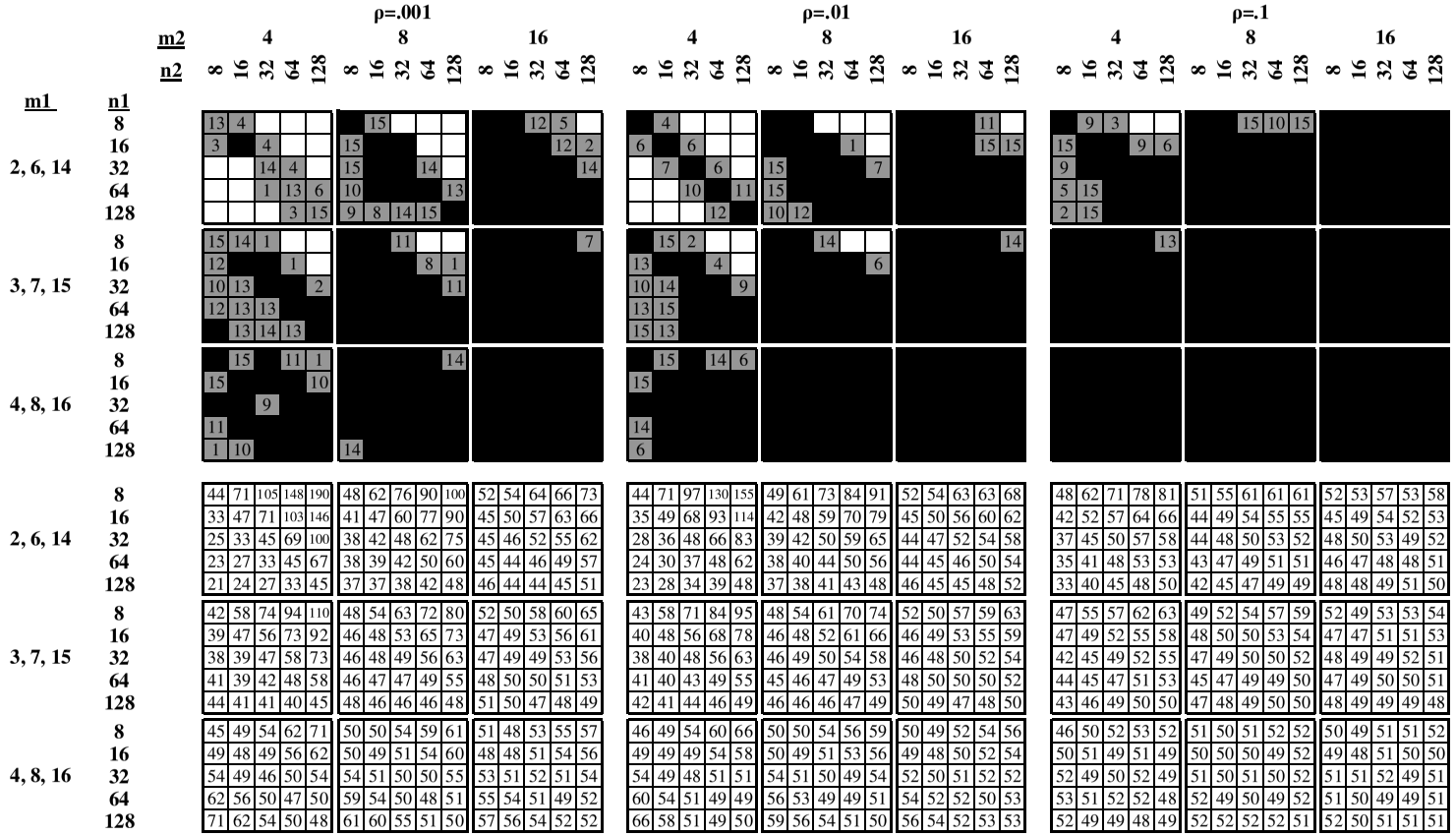
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.4: Type I Error for Test 4 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



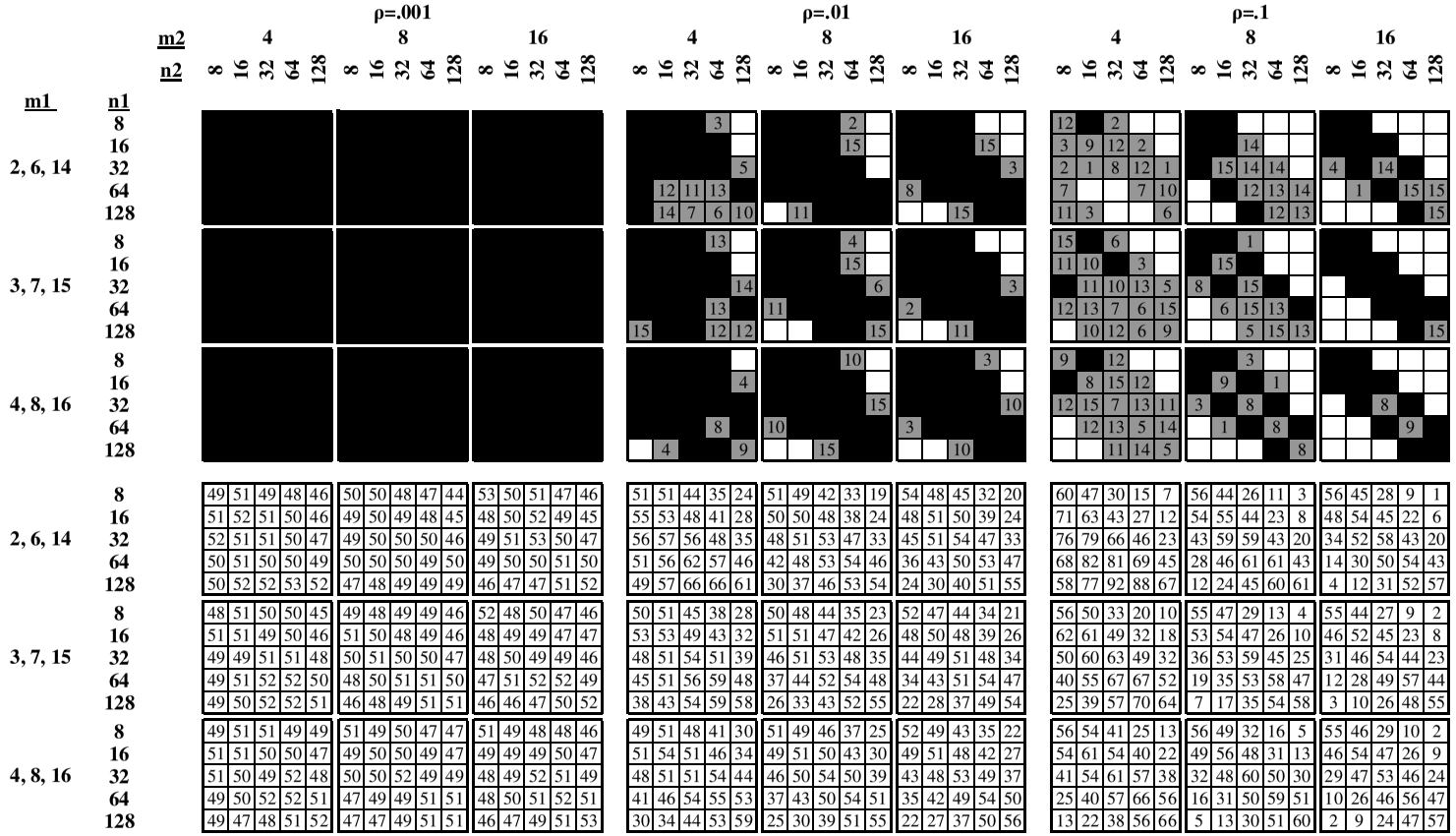
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.5: Type I Error for Test 5 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



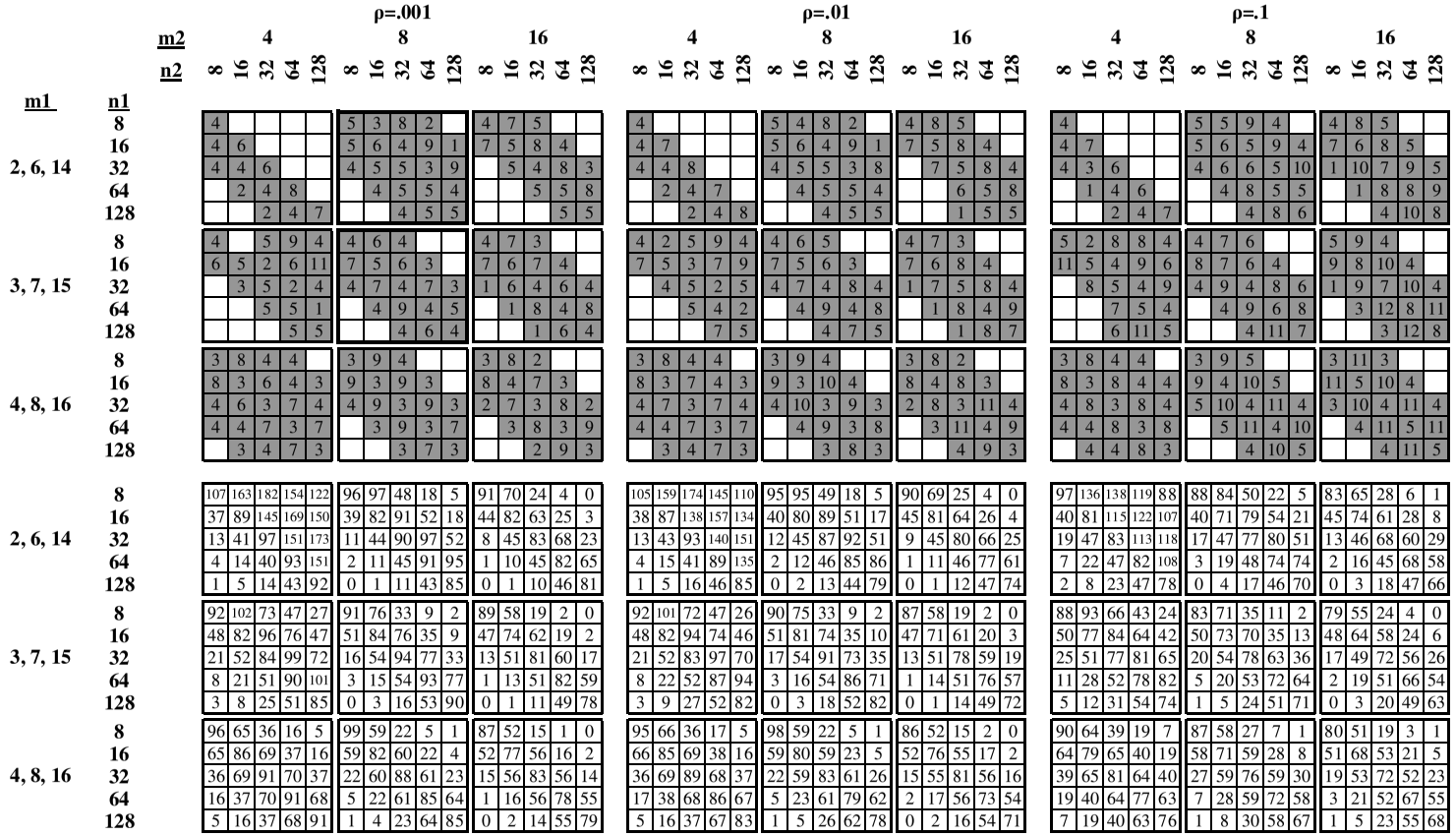
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.6: Type I Error for Test 6 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



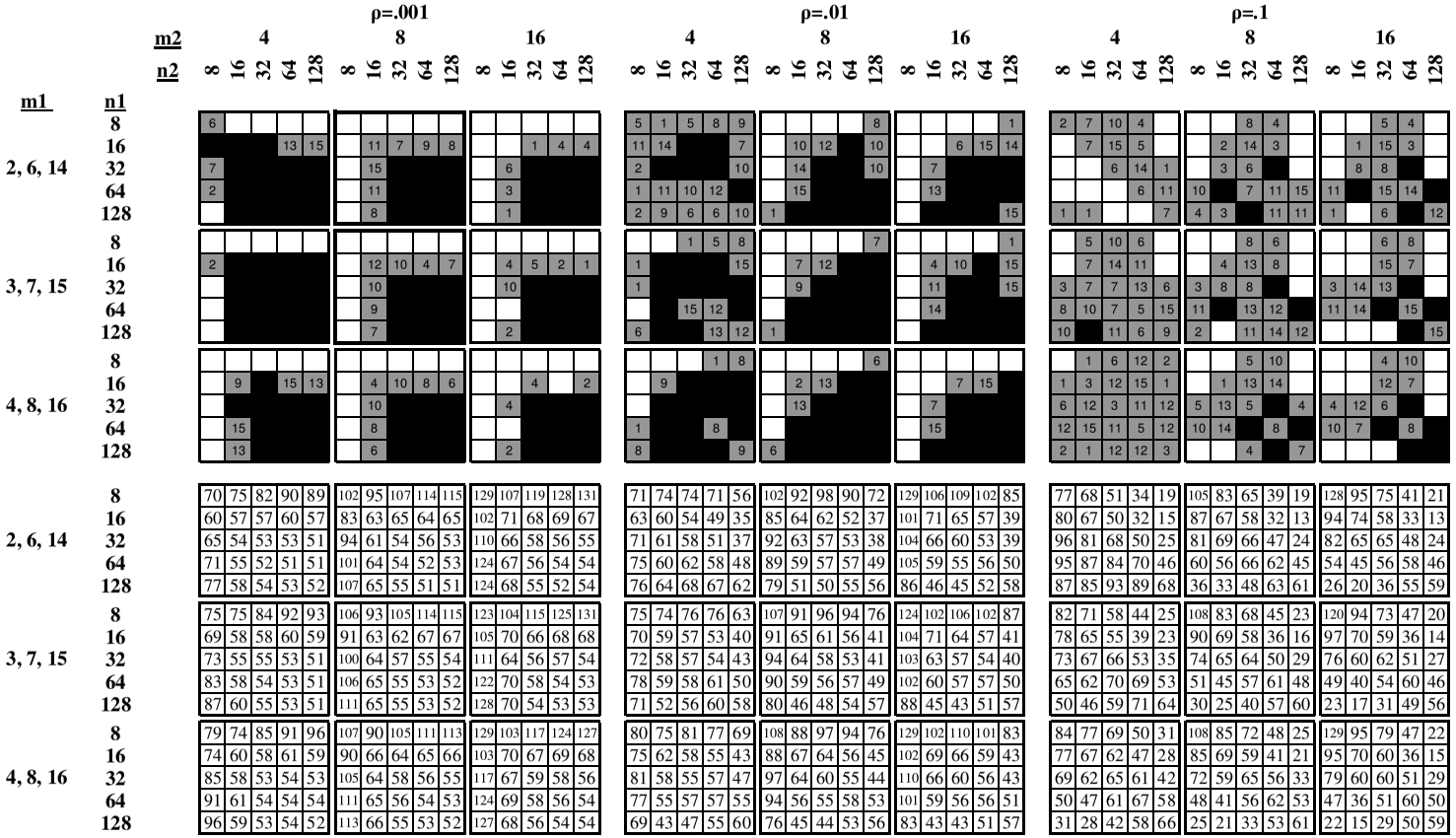
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.7: Type I Error for Test 7 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



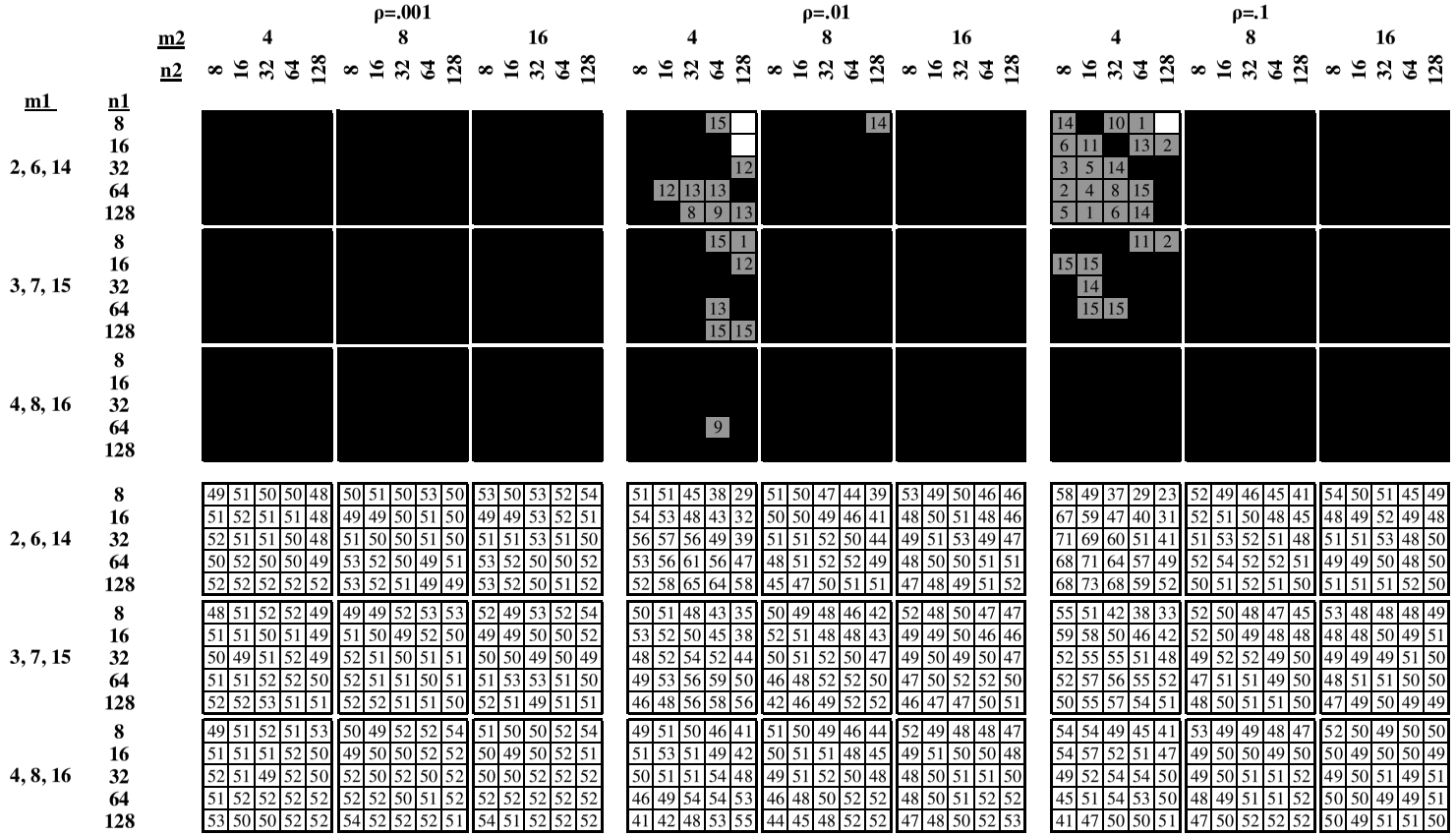
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.8: Type I Error for Test 8 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



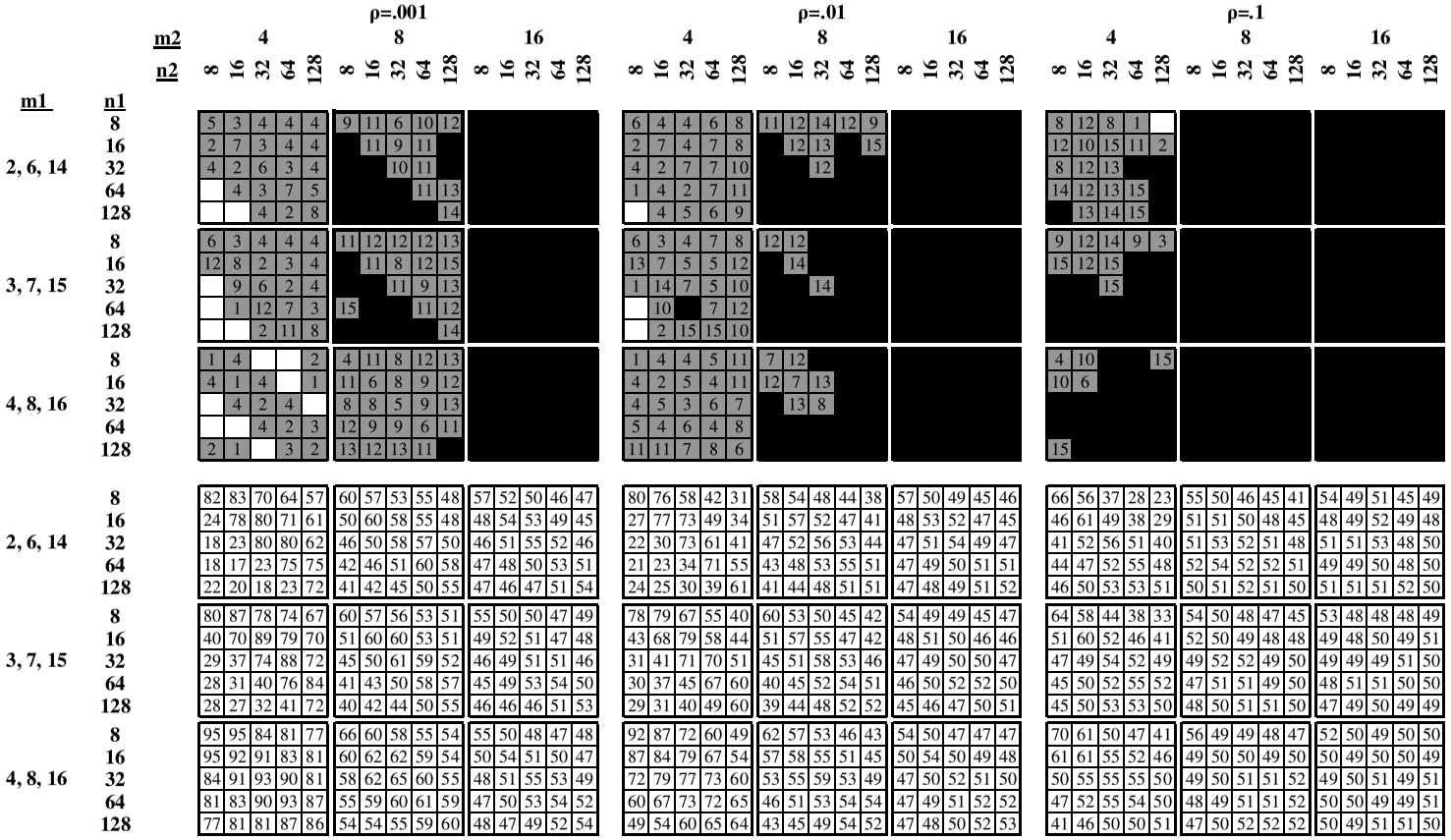
Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.9: Type I Error for Test 9 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Figure 3.10: Type I Error for Test 10 by ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$



Each cell represents 10 (when $m_1 = m_2$ and $\bar{n}_1 = \bar{n}_2$) or 16 (all others) cases of $r_1 \times r_2$ for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$. Black, gray, or white in the first three large rows indicates, respectively, all, some (number indicated) or no, combinations of $r_1 \times r_2$ have unbiased type I error. The second three large rows indicate median type I error for each combination of ρ, m_1, m_2, $\bar{n}_1$, and $\bar{n}_2$.

Chapter 4

Power Analysis for Continuing a Longitudinal Cluster Sample

4.1 Introduction

Simulations of type I error for several one-stage and two-stage analyses of data with one level of clustering were presented in Chapter 3. These concluded that a two-stage analysis of cluster means weighted by the inverse of estimates of their theoretical variance, $\left(\sigma_y^2/n_{hi}\right)\left\{1+\left(n_{hi}-1\right)\rho\right\}$, controls type I error for all but the most extreme cases of imbalance in group randomized trials. Variance components were estimated with restricted maximum likelihood methods in a one-stage model for the individual level data and were constrained to be positive. Other tests considered in Chapter 3 did not control type I error in as many cases or for as wide range of parameters of imbalance, so that this method is preferred over the others evaluated.

A natural extension of the research in Chapter 3 is to evaluate this method with respect to power. Current methods for power computations in group randomized trials focus on data with balanced cluster sizes, for which exact power can be computed. Although studies may be designed using an assumption of balanced data, in practice unbalanced data are ultimately collected and an analysis that employs some type of weighting to account for imbalance is employed. Thus, power for the actual test performed may not agree with power estimated under a simplified version of the data analysis. Methods are needed to perform power analyses for the two stage analysis of means weighted by estimates of their theoretical variance. Such methods would provide power analyses that are aligned with a data analysis method found to perform well (in terms of type I error) under most circumstances for unbalanced clustered data.

Komro *et al.* [16] describes a study to evaluate effects of parental provision of alcohol and home alcohol accessibility on youth alcohol use. The study design was longitudinal within a clustered data framework, as data were collected on students within schools in the 6th, 7th, and 8th grades and schools were randomized to intervention or control conditions. Data were collected on 7 to 86 students in each of 59 schools in Chicago. The study found significant associations between study outcomes and intervention; as such, a continuation of the study for high school grades was desired.

In writing a grant to fund the continuation of this study, a power analysis was required. Because this is a continuation of a previous study, reliable estimates of all required parameters for the power analysis were readily available, including the (unbalanced) number of students per school ultimately sampled as well as covariate averages. Though the data were collected in a longitudinal framework, univariate hypothesis tests were performed. We illustrate how to perform a power analysis for this data using a two-stage analysis of cluster means weighted by estimates of their theoretical variance. We emphasize how this provides a simple way to include individual level covariates into a power analysis for clustered data. We also discuss how these methods of power computation can be extended to multivariate hypotheses.

4.2 Literature Review

4.2.1 Review of Estimation and Hypothesis Testing for Clustered Data

Estimation and hypothesis testing for fixed effects in the analysis of clustered data in the one and two stage models was reviewed in section 3.3. This discussion emphasized that when data have balanced cluster sizes, estimators and hypothesis tests in both models coincide and have known exact distributions and optimal properties. This property is due to the fact that when cluster sizes are balanced, estimators for fixed effects in both the one and two stage models are ordinary not weighted least squares estimators. Thus, they are not functions of (unknown) variance components, and their distribution can be described analytically. Section 3.3 then described the estimators and approximate hypothesis tests for fixed effects when data are unbalanced. It described defensible choices for denominator degrees of freedom for the $\mathcal{F}$ statistic in the one stage model and several choices of weights applied to cluster means in the two stage model. Section 3.6 performed simulations of type I error for various test statistics in

the one and two stage models and concluded that a two-stage analysis of cluster means weighted by the inverse of estimates of their theoretical variance, $\left\{\widehat{\sigma}_y^2\left[1+\left(n_{hi}-1\right)\widehat{\rho}\right]/n_{hi}\right\}^{-1}$, controls type I error for the most cases of imbalance in cluster size, imbalance in number of clusters per treatment, and magnitude of within cluster correlation.

4.2.2 Addition of Covariates

Earlier chapters assumed no covariates other than treatment group were included the models. This assumption was made to remove a layer of complexity in the enumerations and simulations presented in chapters 2 and 3. In practice, data analysis almost always includes covariates other than the effect of interest, either as a means of controlling for potential confounding or to reduce the residual error. When cluster level covariates are included in the one or two stage models, the previous properties of exact and optimal hypothesis tests for data with balanced cluster sizes still hold. This is due to the fact that such data can be shown to meet the assumptions a multivariate linear model, for which exact hypothesis tests for fixed effects exist [23, Ch. 12].

When data include individual level covariates, however, estimators and hypothesis tests for the fixed effects in the one and two stage models differ and neither can be shown to be optimal, even for balanced data. In this case as well as for unbalanced data, estimators in both models can be described as substitution estimators; that is, they perform the test that would be optimal were the variance components known, and estimates of the variance components are inserted in place of desired known quantities. The one stage model substitutes estimates of variance components directly into elements of the covariance matrix; the two stage model substitutes them into weights of the cluster means.

In contrast to the one-stage model for individual level data, covariates enter into the two-stage model simply, as covariate averages. This simplifies understanding of the role individual level covariates play in hypothesis testing for fixed effects of interest, in particular, in computation of power.

4.2.3 Computation of Power for Clustered Data

Earlier chapters focused exclusively on hypothesis testing for fixed effects for the null case, in order to assess type I error properties. Planning for future studies as well as assessment of

the strength of current studies requires computation of test size under the alternative, that is, computation of power.

Current power analyses for clustered trials are performed assuming data have balanced cluster sizes. Murray [24, Ch. 9] and Donner and Klar [7, Ch. 5] gave formulas for computation of power as well as formulas to compute the number of clusters and number of observations per cluster needed to achieve a desired power, given balanced cluster sizes. If unbalance data are expected, users often substitute the arithmetic or harmonic mean number of participants per cluster in these calculations as suggested by Donner and Klar [7]. In addition, Muller *et al.* [22] showed how to compute power for complete compound symmetric data in terms of a multivariate linear model formulation.

Since analyses of clustered data almost always include and compensate for unbalanced cluster sizes, such computations for balanced data are overly simplistic and lead to a misalignment of the power analysis and the actual data analysis performed. Muller *et al.* [22] discussed this issue further for general power computations for Gaussian data. In order to provide an power analysis aligned with the data analysis performed, reliable methods of computing power for unbalanced data are needed.

Muller and Stewart [23, Ch. 26] reviewed methods for computation of power in mixed linear models with any covariance structure, of which the compound symmetry structure of clustered data is a special case. Although some suggestions have been made for mixed model power analyses, necessary for computation of power in the one stage linear model, little is known about the accuracy of these methods. In particular, the current available tests for mixed models have uncertain performance when the number of independent sampling units (for clustered data, the number of clusters) is small [23, Sec. 18.5][20].

In contrast, computing power in the two stage model for cluster means with unbalanced data reduces the problem to computing power for a weighted univariate linear model, for which theory is known and exact given known weights. When variance components are known, optimal weights are equal to the inverse of the theoretical variance of each cluster mean, as described in Chapter 3. Reliable estimates of covariance parameters can be obtained if a previous large study exists. Given these estimates of covariance parameters, computation of power in the two stage analysis of cluster means with weights equal to the inverse of estimates

of the theoretical variance of each cluster mean will have close to exact theory. Further, this model can be transformed to a univariate linear model with homogenous errors. Taylor and Muller [33] showed how to compute confidence limits for power which reflect the uncertainty of power calculations due to estimation of variance components.

4.3 Performing the Power Analysis

4.3.1 Data Structure

We consider data with one level of clustering, so that individual observations are nested within cluster. In the following notation, clusters are also nested within treatment group. Let $h = 1, \ldots, g$ index treatment groups, $i = 1, \ldots, m_h$ index clusters within treatment group h, and let $j = 1, \ldots, n_{hi}$ index observations within treatment cluster i and treatment group h. Denote the total number of clusters by $m = \sum_{h=1}^{g} m_h$, the number of observations in treatment group h by $n_h = \sum_{i=1}^{m_h} n_{hi}$, and the total number of observations by $N = \sum_{h=1}^{g} \sum_{i=1}^{m_h} n_{hi}$.

4.3.2 Statement of One-Stage Model

Define a linear model for continuous Gaussian outcome $\boldsymbol{y}_1$ that includes fixed effects given in $\boldsymbol{\beta}$ $(q \times 1)$, a random effect for cluster given in $\boldsymbol{b}$ $(m \times 1)$, and a random error, $\boldsymbol{e}_1$ $(N \times 1)$:

$$\boldsymbol{y}_1 = \boldsymbol{X}_1\boldsymbol{\beta} + \boldsymbol{Z}_1\boldsymbol{b} + \boldsymbol{e}_1. \tag{4.1}$$

The matrices $\boldsymbol{X}_1$ $(N \times q)$ and $\boldsymbol{Z}_1$ $(N \times m)$ are design matrices for the fixed and random effects, respectively.

Vectors or matrices $\boldsymbol{y}_1$, $\boldsymbol{X}_1$, $\boldsymbol{Z}_1$,and $\boldsymbol{e}_1$ are stacked by treatment group and cluster so that $\boldsymbol{y}_1 = \left\{{}_c\, \boldsymbol{y}_{1,hi}\right\}$, $\boldsymbol{X}_1 = \left\{{}_c\, \boldsymbol{X}_{1,hi}\right\}$, $\boldsymbol{Z}_1 = \left\{{}_c\, \boldsymbol{Z}_{1,hi}\right\}$, and $\boldsymbol{e}_1 = \left\{{}_c\, \boldsymbol{e}_{1,hi}\right\}$. This chapter discusses analyses with covariates other that the main variable of interest; previous chapters did not. As such, define $\boldsymbol{X}_1 = \boldsymbol{X}_{1t}||\boldsymbol{X}_{1c}$, where $\boldsymbol{X}_{1t}$ and $\boldsymbol{X}_{1c}$ are the fixed effects design matrices for main effects parameters of interest and all other covariates, respectively. The design matrix for the random cluster effect (the only random effect) is $\boldsymbol{Z}_1 = \left\{{}_d\, \mathbf{1}_{n_{hi}}\right\}$.

We assume $\boldsymbol{b} \sim \mathcal{N}_m\left(\mathbf{0}, \sigma_c^2 \boldsymbol{I}_m\right)$ independently of $\boldsymbol{e}_1 \sim \mathcal{N}_N\left(\mathbf{0}, \sigma_e^2 \boldsymbol{I}_N\right)$ so that:

$$\boldsymbol{y}_1 \sim \mathcal{N}_N\left(\boldsymbol{X}_1\boldsymbol{\beta}, \boldsymbol{\Sigma}_1\right),$$

where the covariance matrix $\boldsymbol{\Sigma}_1$ $(N \times N)$ is compound symmetric and has the form:

$$\boldsymbol{\Sigma}_1 = \sigma_c^2 \boldsymbol{Z}_1 \boldsymbol{Z}_1' + \sigma_e^2 \boldsymbol{I}_N = \left\{ {}_d\, \sigma_c^2 \mathbf{1}_{n_{hi}} \mathbf{1}_{n_{hi}}' + \sigma_e^2 \boldsymbol{I}_{n_{hi}} \right\}.$$

$\boldsymbol{\Sigma}_1$ may be expressed in terms of the total variance, σ_y^2, and within cluster correlation, ρ, as:

$$\boldsymbol{\Sigma}_1 = \sigma_y^2 \left\{ {}_d\, \boldsymbol{I}_{n_{hi}} + \left(\mathbf{1}_{n_{hi}} \mathbf{1}_{n_{hi}}' - \boldsymbol{I}_{n_{hi}} \right) \rho \right\},$$

where $\rho = \sigma_c^2 / \left(\sigma_c^2 + \sigma_e^2 \right)$ and $\sigma_y^2 = \sigma_c^2 + \sigma_e^2$ or, equivalently, $\sigma_c^2 = \sigma_y^2 \rho$ and $\sigma_e^2 = \sigma_y^2 (1 - \rho)$. Implicit in construction of $\boldsymbol{\Sigma}_1$ is the assumption that data across all treatment groups have the same variance parameters.

4.3.3 Statement of Two-Stage Model

To transform from a model for individual level data to a model for cluster means, pre-multiply model (3.1) by the matrix $\boldsymbol{T}_1$ $(m \times N)$, where $\boldsymbol{T}_1 = \left\{ {}_d\, \mathbf{1}_{n_{hi}}' / n_{hi} \right\}$. This yields a model for $\boldsymbol{y}_2$ $(m \times 1) = \boldsymbol{T}_1 \boldsymbol{y}_1$ where:

$$\boldsymbol{y}_2 = \boldsymbol{X}_2 \boldsymbol{\beta} + \boldsymbol{Z}_2 \boldsymbol{b} + \boldsymbol{e}_2, \tag{4.2}$$

and $\boldsymbol{X}_2$ $(m \times q) = \boldsymbol{T}_1 \boldsymbol{X}_1$, $\boldsymbol{Z}_2$ $(m \times m) = \boldsymbol{T}_1 \boldsymbol{Z}_1$, and $\boldsymbol{e}_2$ $(m \times 1) = \boldsymbol{T}_1 \boldsymbol{e}_1$. Parameters in $\boldsymbol{\beta}$ and $\boldsymbol{b}$ were not affected by the transformation.

The vector of outcomes, $\boldsymbol{y}_2$, and of random errors, $\boldsymbol{e}_2$, contain cluster averages, so that $\boldsymbol{y}_2 = \left\{ {}_c\, \overline{y}_{hi} \right\}$ and $\boldsymbol{e}_2 = \left\{ {}_c\, \overline{e}_{hi} \right\}$. In line with notation for the one-stage model, define $\boldsymbol{X}_2 = \boldsymbol{X}_{2t} || \boldsymbol{X}_{2c}$, where $\boldsymbol{X}_{2t} = \boldsymbol{T}_1 \boldsymbol{X}_{1t}$ and $\boldsymbol{X}_{2c} = \boldsymbol{T}_1 \boldsymbol{X}_{1c}$ are the fixed effects design matrices for main effects parameters of interest and all other covariates, respectively. $\boldsymbol{X}_{2c}$ contains covariate cluster averages. For example, when $\boldsymbol{X}_{2c}$ contains only one covariate, $\boldsymbol{x}$, $\boldsymbol{X}_{2c} = \left\{ {}_c\, \overline{x}_{hi} \right\}$. The design matrix for the random effects has form $\boldsymbol{Z}_2 = \boldsymbol{I}_m$.

In line with previous assumptions, we assume $\boldsymbol{b} \sim \mathcal{N}_m \left(\mathbf{0}, \sigma_c^2 \boldsymbol{I}_m \right)$ independently of $\boldsymbol{e}_2 \sim \mathcal{N}_m \left(\mathbf{0}, \sigma_e^2 \boldsymbol{T}_1 \boldsymbol{T}_1' \right)$ so that:

$$\boldsymbol{y}_2 \sim \mathcal{N}_m \left(\boldsymbol{X}_2 \boldsymbol{\beta}, \boldsymbol{\Sigma}_2 \right),$$

where $\boldsymbol{\Sigma}_2$ $(m \times m)$ is given by:

$$\boldsymbol{\Sigma}_2 = \boldsymbol{T}_1 \boldsymbol{\Sigma}_1 \boldsymbol{T}_1' = \left\{ {}_d\, \sigma_c^2 + \sigma_e^2 / n_{hi} \right\}.$$

In terms of the alternate parameterization with $\left(\sigma_y^2, \rho\right)$ instead of $\left(\sigma_e^2, \sigma_c^2\right)$:

$$\boldsymbol{\Sigma}_2 = \sigma_y^2 \big\{_d \left[1 + (n_{hi} - 1)\,\rho\right] / n_{hi} \big\}.$$

4.3.4 General Linear Hypothesis

Define a vector of secondary contrast parameters, $\boldsymbol{\theta}\ (a \times 1) = \boldsymbol{C\beta}$, where $\boldsymbol{C}\ (a \times g)$ contains desired contrasts for the fixed effects. We study the two sided general linear hypothesis (GLH) $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. In most hypotheses of interest, $\boldsymbol{\theta}_0 = \mathbf{0}$. Such a hypothesis test describes differences in the fixed effects only. The $\boldsymbol{X}$ matrices we consider are full rank, ensuring $\boldsymbol{\theta}$ is estimable. Requiring $\boldsymbol{C}$ to have full row rank [$\text{rank}(\boldsymbol{C}) = a$] ensures the GLH is testable [21].

4.3.5 Theory for Power Computation

Consider the two-stage model for $\boldsymbol{y}_2$ given in equation 4.2. For unbalanced clustered data, observations in $\boldsymbol{y}_2$ are independent with heterogeneous variances. As discussed earlier in sections 2.2.4 and 3.4.2, exact estimation and hypothesis testing for fixed effects can be obtained under both the null and alternative hypotheses when $\boldsymbol{\Sigma}_2$ can be written in the form $\boldsymbol{\Sigma}_2 = \sigma^2 \boldsymbol{W}^{-1}$, where $\boldsymbol{W}$ is a known and constant matrix. In this case, software such as PROC GLMPOWER in SAS could be used to compute power for this weighted univariate linear model. Optimal weights are the inverse of the theoretical variance of the cluster means, $\boldsymbol{W} = \left\{\left[1 + (n_{hi} - 1)\,\rho\right] / n_{hi}\right\}^{-1}$. The multiplier for σ_y^2 is not needed in the weights, since it is constant for all clusters and cancels out of calculations.

In this section, we employ an additional transformation so that elements of $\boldsymbol{y}_2$ are independent with homogenous rather than heterogeneous variances and therefore satisfy the assumptions of the general linear univariate model (GLUM). Muller and Fetterman [21] and Muller and Stewart [23, Ch. 2] among others, give extensive information about this model. While few software packages compute power for the weighted univariate linear model, most statistical software packages include computations for power in the GLUM. We perform power calculations with a suite of SAS/IML modules, POWERLIB [10]. Additionally, Taylor and Muller [33] showed how to compute exact confidence intervals for power in the GLUM that reflect uncertainty introduced with estimation of variance parameters, here ρ.

To transform to a model with independent observations with homogenous variances, pre-multiply the model 4.2 by the matrix $\boldsymbol{T}_2 = \left\{_d \left(n_{hi} / \left[1 + \left(n_{hi} - 1\right)\rho\right]\right)^{-1/2}\right\}$. This yields a model for $\boldsymbol{y}_3\ (m \times 1) = \boldsymbol{T}_2\boldsymbol{y}_2$:

$$\boldsymbol{y}_3 = \boldsymbol{X}_3\boldsymbol{\beta} + \boldsymbol{Z}_3\boldsymbol{b} + \boldsymbol{e}_3, \tag{4.3}$$

where $\boldsymbol{X}_3\ (m \times g) = \boldsymbol{T}_2\boldsymbol{X}_2$, $\boldsymbol{Z}_3\ (m \times m) = \boldsymbol{T}_2\boldsymbol{Z}_2$, and $\boldsymbol{e}_3\ (m \times 1) = \boldsymbol{T}_2\boldsymbol{e}_2$. Parameters in $\boldsymbol{\beta}$ and $\boldsymbol{b}$ were not affected by the transformation.

In line with previous assumptions, we assume $\boldsymbol{b} \sim \mathcal{N}_m\left(\boldsymbol{0}, \sigma_c^2\boldsymbol{I}_m\right)$ independently of $\boldsymbol{e}_3 \sim \mathcal{N}_m\left(\boldsymbol{0}, \sigma_e^2\boldsymbol{T}_2\boldsymbol{T}_1\boldsymbol{T}_1'\boldsymbol{T}_2'\right)$ so that:

$$\boldsymbol{y}_3 \sim \mathcal{N}_m\left(\boldsymbol{X}_3\boldsymbol{\beta}, \boldsymbol{\Sigma}_3\right),$$

where $\boldsymbol{\Sigma}_3\ (m \times m)$ is given by:

$$\boldsymbol{\Sigma}_3 = \boldsymbol{T}_2\boldsymbol{\Sigma}_2\boldsymbol{T}_2' = \sigma_y^2\boldsymbol{I}_m.$$

Following theory described in Muller and Fetterman [21, Ch. 17], power for a hypothesis test for the fixed effects $\boldsymbol{\beta}$ can then be computed as:

$$\text{Power} = 1 - \text{Prob}\left(\mathcal{F}_{a,m-r,\omega} < f_c\right),$$

where $r = \text{rank}\left(\boldsymbol{X}_3\right)$, $f_c = F_{\mathcal{F}}^{-1}\left(1 - \alpha, a, m - r\right)$, and the noncentrality, ω is:

$$\omega = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)'\left[\boldsymbol{C}\left(\boldsymbol{X}_3'\boldsymbol{X}_3\right)\boldsymbol{C}'\right]^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)/\sigma_y^2.$$

4.3.6 Steps to Perform Power Analysis

The power analysis described in the previous section may be computed with the following steps:

1. Specify α, $\boldsymbol{C}$, $\boldsymbol{\theta}_0$, and cluster sizes.
2. Obtain ρ from a previous study. In most cases, σ_c^2 and σ_e^2 will be estimated from a one level model for individual level data, then $\rho = \sigma_c^2/\left(\sigma_c^2 + \sigma_e^2\right)$ is computed.
3. Compute $\boldsymbol{T}_1 = \left\{_d \boldsymbol{1}_{n_{hi}}'/n_{hi}\right\}$.
4. Obtain $\boldsymbol{X}_1$ from a previous study and compute $\boldsymbol{X}_2 = \boldsymbol{T}_1\boldsymbol{X}_1$, or use $\boldsymbol{X}_2$ from a previous study.
5. Compute $\boldsymbol{T}_2 = \left\{_d \left(n_{hi} / \left[1 + \left(n_{hi} - 1\right)\rho\right]\right)^{-1/2}\right\}$.

6. Compute $\boldsymbol{y}_3 = \boldsymbol{T}_2\boldsymbol{y}_2$, $\boldsymbol{X}_3 = \boldsymbol{T}_2\boldsymbol{X}_2$, and $r = \text{rank}\,(\boldsymbol{X}_3)$.

7. Compute $\sigma^2 = \boldsymbol{y}_3' \left[\boldsymbol{I}_m - \boldsymbol{X}_3 \left(\boldsymbol{X}_3'\boldsymbol{X}_3\right)^{-1} \boldsymbol{X}_3'\right] \boldsymbol{y}_3$.

8. In order to obtain parameter values for covariates, compute $\boldsymbol{\beta} = \left(\boldsymbol{X}_3'\boldsymbol{X}_3\right)^{-1} \boldsymbol{X}_3'\boldsymbol{y}_3$. This has the form $\boldsymbol{\beta} = \boldsymbol{\beta}_t // \boldsymbol{\beta}_c$ where $\boldsymbol{\beta}_t$ is the effect of interest and $\boldsymbol{\beta}_c$ are covariates. Replace $\boldsymbol{\beta}_t$ with desired mean difference parameter(s).

9. Compute the noncentrality, $\omega = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \left[\boldsymbol{C}\left(\boldsymbol{X}_3'\boldsymbol{X}_3\right)\boldsymbol{C}'\right]^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) / \sigma_y^2$.

10. Compute the critical value $f_c = F_{\mathcal{F}}^{-1}(1 - \alpha, a, m - r)$.

11. Compute Power $= 1 - \text{Prob}\,(\mathcal{F}_{a,m-r,\omega} < f_c)$.

4.4 Data Example

Komro *et al.* [16] described a study to evaluate effects of parental provision of alcohol and home alcohol accessibility on youth alcohol use. The data was collected from 1,388 students, and their parents, who attended Chicago public schools that were randomized to the intervention or control group of a previous group randomized trial to study alcohol prevention. Students completed self report questionnaires at the beginning of 6th grade (considered baseline) and at the end of the 6th, 7th, and 8th grades with response rates ranging from 91% to 96%. These questionnaires assessed the availability of alcohol at home, student alcohol use, and alcohol use intentions. Parents also completed questionnaires when their child started 6th grade. The parent questionnaires assessed family communication, alcohol practices, and beliefs about youth alcohol use. Due to repeated measures on each student at four time points as well as correlation of students within schools, these data can be said to have a block kronecker compound symmetric by unstructured covariance structure within each school. The study showed several significant relationships between youth alcohol use and both of parental provision of alcohol and home alcohol accessibility. As such, a continuation of this study into high school years (grades 9 - 12) was desired. The following describes the power analyses performed in the grant to obtain funding for the continuation study.

Four univariate outcomes were considered; a similar power analysis was performed for each. The following describes the power analysis for one of these outcomes, change in a continuous

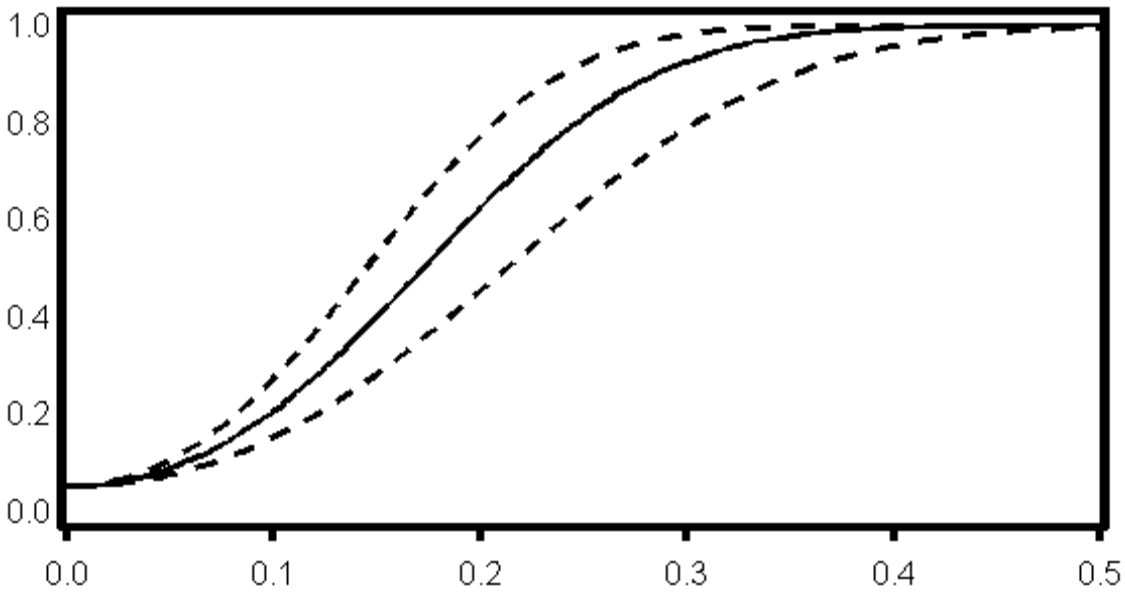
measure of student alcohol use from 9th grade to 12th grade. This outcome was obtained by summing and scaling student responses to several items in the alcohol use questionnaire in the 9th and 12th grades and subtracting their values. Predictors in the model included number of alcohol ads seen, parental communication styles, alcohol access in the home, parental monitoring, and alcohol norms. The analysis desired was to study whether magnitude of alcohol norms predicted change in alcohol use from the 9th to 12th grades.

For this study, $\boldsymbol{C} = 1$, $\alpha = 0.05$, and $\boldsymbol{\theta}_0 = 0$. The observed cluster sizes ranged from 7 to 86 with mean and median equal to 32. The design matrix $\boldsymbol{X}_1$ as well as an estimate of ρ were obtained from the study on middle school students. Power was computed following the steps outlined in section 4.3.6. Figure 4.1 shows power for various values of mean difference with confidence limits to reflect uncertainty in power calculations due to estimation of ρ.

4.5 Further Remarks

This chapter described how to perform a univariate power analysis for unbalanced clustered data in the two stage model for cluster means weighted by the inverse of their theoretical variance. This method gives a power analysis for unbalanced clustered data that matches a defensible data analysis. A strength of this method is the simple and valid way in which it incorporates individual level covariates into the power analysis as covariates averages. This method considered only a univariate outcome; future research will explore how to extend this method to multivariate outcomes. Such multivariate outcomes have a Kronecker compound symmetry by unstructured covariance matrix within each (independent) cluster. Extension of this method to multivariate outcomes should be straightforward, because the transformation to cluster means removes the compound symmetry layer of the covariance structure and leaves in almost all cases, complete unstructured data.

Figure 4.1: Power as a Function Mean Difference



Chapter 5

Summary and Future Research

Due to logistical, cost, or ethical constraints as well as to purposely study relationships of variables on a community level, public health studies often employ a clustered data collection design instead of a fully randomized design. As such, evaluations of current estimation and hypothesis testing procedures for clustered data, as well as suggestions of new and better methods, are desired. This dissertation focused on evaluating methods of estimation and hypothesis testing for fixed effects in a univariate or repeated measures mixed linear model with Gaussian errors and one level of clustering. This dissertation also suggested a better way to do power analysis for clustered data than is currently being performed.

Because of the equal correlation structure of observations with clusters, called compound symmetry, clustered data have the special property that equivalent inference for cluster level variables may be made with either an analysis of cluster means or an analysis of individual level data when data have balanced clustered sizes. Estimation and hypothesis testing for fixed effects differs for the two analyses when data have unbalanced cluster sizes. The analysis of cluster means fits a weighted univariate linear model. Such a model computes ordinary least squares estimates for fixed effects and their associated hypothesis tests; also, estimates of variance components are computed non-iteratively. The analysis of individual level data is performed via a random effects model, where the random effect is due to cluster. This model computes approximate weighted least squares estimators for fixed effects and their associated Wald-style hypothesis tests. Estimation of variance components in the analysis of individual level data usually must be conducted with iterative procedures.

Results of an enumeration study of type I error for the analysis of cluster means with means

unweighted or weighted by cluster size were presented in Chapter 2. Scenarios of imbalance common to non-randomized studies were considered; in particular, these included scenarios of imbalance in the number of clusters per treatment group. These enumerations showed that when treatment groups have an equal number of clusters per group, an analysis of unweighted means has unbiased type I error. In turn, when the within cluster correlation is small, weighting by cluster size also has unbiased type I error. Type I errors for hypothesis testing with both types of weights were unbiased when average cluster size was equal for the two treatment groups. One surprising finding was that the magnitude of ratio of maximum to minimum cluster size per treatment group affected type I error relatively little. Future research will extend this enumerations study to other hypotheses (e.g., more treatment groups), addition of covariates, and additional levels of clustering. This enumeration study of type I error could also be replicated for power for cases of imbalance in which type I error was unbiased.

Also presented in Chapter 2 was a theoretical method that can be used to compute exact probabilities from the distribution of the two stage model test statistic for any known weights when variance components are known. Such theory made enumerations of probabilities possible rather than requiring simulations. The theory is applicable to data with any covariance structure, not just the diagonal but heterogeneous covariance structure of cluster means, though that application is highlighted in this dissertation. Future research will explore use of this method for data with other covariance structures.

Simulations of type I error for ten hypothesis tests for clustered data were presented in Chapter 3. These included several methods of analysis of individual level data and several analyses of weighted cluster means. The simulations concluded that a hybrid approach of the two stage model with weights equal to the inverse of estimates of the theoretical variance of the cluster means controlled type I error for more cases than other tests. In this method, variance components were estimated from the individual level data. The test which controlled type I error for the closest number of cases to the previous method was the one stage analysis with Kenward and Roger [13] degrees of freedom, where variance components were either constrained or not constrained to be positive. The model with constrained estimates of variance components was undesirable because for this method, type I error was most biased for cases close to balance. The model with unconstrained variance components controlled type I error for balanced cases,

but often did not converge when the variance components would have been estimated to be negative. Findings from Chapter 2 about the analysis of unweighted means and means weighted by cluster size were also replicated. Future research will compare these 10 tests with respect to power, for cases of imbalance in which type I error was unbiased.

Chapter 4 showed how to perform a power analysis for the two stage method that controlled type I error well in Chapter 3. Current power analyses for clustered data assume balanced data; this power analysis gives a way to compute power for unbalanced data. Studies usually obtain unbalanced data, so this method gives a power analysis aligned with a defensible data analysis for unbalanced clustered data. This method also easily incorporates individual level covariates into power calculations via covariate averages. Further, this method for computing power naturally extends to repeated measures clustered data designs. If data have compound symmetric kronecker unstructured covariance within each school, transforming the cluster means reduces the covariance pattern by one level and leads to cluster means with an unstructured covariance. Such data will almost always be complete and balanced, so that small sample multivariate power methods can be used to compute power. This is a dramatic improvement on previous methods, as no reliable methods for computing power for multivariate clustered data have been described before. Future research will explore use of this technique for multivariate data.

An area of future application that was not discussed in this dissertation is derivation of theory for internal pilot studies with clustered data. These studies estimate covariance parameters in the middle of data collection procedures and increase or decrease final sample size based on these interim estimates. Current research has shown how to perform internal pilots calculations for complete balanced compound symmetric data. A natural extension of this is to theory for internal pilots with compound symmetric data with missing observations such as in unbalanced clustered data.

Bibliography

[1] Box GEP (1954). "Some Theorems of Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification." *Annals of Mathematical Statistics*, **25**(2), 290–302.

[2] Bryk AS, Raudenbush SW (1992). *Hierarchical Linear Models*. Sage Publications, Newbury Park.

[3] Catellier DJ, Muller KE (2000). "Tests for Gaussian Repeated Measures with Missing Data in Small Samples." *Statistics in Medicine*, **19**, 1101–1114.

[4] Davies RB (1980). "Algorithm AS 155: The Distribution of a Linear Combination of Chi-Square Random Variables." *Applied Statistics*, **29**(3), 323–333.

[5] Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood From Incomplete Data Via EM Algorithm." *Journal Of The Royal Statistical Society Series B-Methodological*, **39**(1), 1–38.

[6] Dempster AP, Rubin DB, Tsutakawa RK (1981). "Estimation in Covariance Components Models." *Journal Of The American Statistical Association*, **76**(374), 341–353.

[7] Donner A, Klar N (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold, New York.

[8] Goldstein H (1987). *Multilevel Models in Educational and Social Research*. Oxford University Press, New York.

[9] Hopkins K (1982). "The Unit of Analysis: Group Means Versus Individual Observations." *American Educational Research Journal*, **19**(1), 5–18.

[10] Johnson JL, Muller KE, Slaughter JC, Gurka MJ, Gribbin MJ, Simpson SL (2007). "POWERLIB: SAS/IML Software for Computing Power in Multivariate Linear Models." *Journal of Statistical Software*, **Under review**.

[11] Johnson NL, Kotz S (1970). *Continuous Univariate Distributions - 2*. Wiley, New York.

[12] Kacker RN, Harville DA (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models." *Journal Of The American Statistical Association*, **79**(388), 853–862.

[13] Kenward MG, Roger JH (1997). "Small sample inference for fixed effects from restricted maximum likelihood." *Biometrics*, **53**(3), 983–997.

[14] Khuri AI, Mathew T, Sinha BK (1998). *Statistical Tests for Mixed Linear Models*. Wiley, New York.

[15] Kim HY, Gribbin MJ, Muller KE, Taylor DJ (2006). "Analytic, computational, and approximate forms for ratios of noncentral and central Gaussian quadratic forms." *Journal Of Computational And Graphical Statistics*, **15**(2), 443–459.

[16] Komro KA, Maldonado-Molina MM, Tobler AL, Bonds JR, Muller KE (2007). "Effects of Home Access and Availability of Alcohol on Young Adolescents Alcohol Use." *Addiction*, **In press**.

[17] Kotz S, Johnson NL, Balakrishnan N (2000). *Continuous Multivariate Distributions, Models and Applications, Vol. 1.* Wiley, New York.

[18] Littell R, Milliken GA, Stroup WW, Wolfinger RD (1996). *SAS System for Mixed Models.* SAS Institute Inc., Cary, NC.

[19] McCulloch CE, Searle SR (2001). *Generalized, Linear, and Mixed Models.* Wiley, New York.

[20] Muller KE, Edwards LJ, Simpson SL, Taylor DJ (2007). "Accurate test size and power in small samples of repeated measures without missing data." *Statistics in Medicine*, **In press**.

[21] Muller KE, Fetterman BA (2002). *Regression and ANOVA: An Integrated Approach to Using SAS Sofware.* SAS Institute Inc., Cary, NC.

[22] Muller KE, Lavange LM, Ramey SL, Ramey CT (1992). "Power Calculations For General Linear Multivariate Models Including Repeated Measures Applications." *Journal Of The American Statistical Association*, **87**(420), 1209–1226.

[23] Muller KE, Stewart PW (2006). *Linear Model Theory: Univariate, Multivariate, and Mixed Models.* Wiley, Hoboken, NJ.

[24] Murray DM (1998). *Design and Analysis of Group-Randomized Trials.* Oxford University Press, New York.

[25] Murray DM, Catellier DJ, Hannan PJ, Treuth MS, Stevens J, Schmitz KH, Rice JC, Conway TL (2004). "School-level intraclass correlation for physical activity in adolescent girls." *Medicine And Science In Sports And Exercise*, **36**(5), 876–882.

[26] Puntanen S, Styan GPH (1989). "The Equality Of The Ordinary Least-Squares Estimator And The Best Linear Unbiased Estimator." *American Statistician*, **43**(3), 153–161.

[27] Sahai H, Ageel M (2000). *The Analysis of Variance: Fixed, Random, and Mixed Models.* Birkhauser, Boston.

[28] Satterthwaite FE (1941). "Synthesis of Variance." *Psychometrika*, **6**(5), 309–316.

[29] Schaalje GB, McBride JB, Fellingham GW (2002). "Adequacy of approximations to distributions of test statistics in complex mixed linear models." *Journal Of Agricultural Biological And Environmental Statistics*, **7**(4), 512–524.

[30] Schott JR (1997). *Matrix Analysis for Statistics.* Wiley, New York.

[31] Searle SR (1971). *Linear Models.* Wiley, New York.

[32] Stevens J, Murray DM, Catellier DJ, Hannan PJ, Lytele LA, Elder JP, Young DR, Simons-Morton DG, Webber LS (2005). "Design of the trial of activity in adolescent girls (TAAG)." *Contemporary Clinical Trials*, **26**(2), 223–233.

[33] Taylor DJ, Muller KE (1995). "Computing Confidence-Bounds For Power And Sample-Size Of The General Linear Univariate Model." *American Statistician*, **49**(1), 43–47.

[34] Tian YG, Wiens DP (2006). "On equality and proportionality of ordinary least squares, weighted least squares and best linear unbiased estimators in the general linear model." *Statistics and Probability Letters*, **76**(12), 1265–1272.

[35] Varnell SP, Murray DM, Janega JB, Blitstein JL (2004). "Design and analysis of group-randomized trials: A review of recent practices." *American Journal Of Public Health*, **94**(3), 393–399.

[36] Verbeke G, Molenberghs G (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.