AN ABSTRACT OF THE THESIS OF

Byung S. Park for the degree of Doctor of Philosophy in Statistics presented on January 3, 2002. Title: Testing Hypotheses using Unweighted Means

Abstract approved: Redacted for Privacy

Justus F. Seely

Testing main effects and interaction effects in factorial designs are basic content in statistics textbooks and widely used in various fields. In balanced designs there is general agreement on the appropriate main effect and interaction sums of squares and these are typically displayed in an analysis of variance (ANOVA). A number of methods for analyzing unbalanced designs have been developed, but in general they do not lead to unique results. For example, in SAS one can get three different main effect sums of squares in an unbalanced design. I, If these results are viewed from the theory of the general linear model, then it is typically the case that the different sums of squares all lead to F-tests, but they are testing different linear hypotheses. In particular, if one clearly specifies the linear hypothesis being tested, then linear model theory leads to one unique deviation sum of squares. One exception to this statement is an ANOVA, called an unweighted means ANOVA (UANOVA) introduced by Yates (1934). The deviation sum of squares in a UNANOVA typically does not lead to an F-test and hence does not reduce to a usual deviation sum of squares for some linear hypothesis.

The UANOVA tests have been suggested by several writers as an alternative to the usual tests. Almost all of these results are limited to the one-way model or a two-way model with interaction, and hence the UANOVA procedure is not available for a general linear model. This thesis generalizes the UANOVA test prescribed in the two-way model with interaction to a general linear model. In addition, the properties of the test are investigated. It is found, similar to the usual test, that computation of the UANOVA test statistic does not depend on how the linear hypothesis is formulated. It is also shown that the numerator of the UANOVA test is like a linear combination of independent chi-squared random variables as opposed to a single chi-squared random variable in the usual test. In addition we show how the Imhof (1961) paper can be used to determine critical values, p-values and power for the UANOVA test. Comparisons with the usual test are also included. It is found that neither test is more powerful than the other. Even so, for most circumstances our general recommendation is that the usual test is probably superior to the UANOVA test.

Testing Hypotheses using Unweighted Means

by

Byung S. Park

A THESIS

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented January 3, 2002

Commencement June 2002

Doctor of Philosophy thesis of Byung S. Park presented January 3, 2002

APPROVED:

Redacted for Privacy

Major Professor, representing Statistics

Redacted for Privacy

Chair of Department of Statistics

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Byung S. Park, Author

ACKNOWLEDGMENT

Many people have contributed substantially to my education, this dissertation and the completion of my degree. I am deeply indebted to Dr. Justus Seely. Without his patience, generosity and support, I could not complete this degree. He is always more than a teacher to me, sometimes I mistaken him for father. I would like to thank to Drs. David Birkes and Youngjo Lee for their useful recommendations and comments. I would like to thank to Dr. David Butler for his supportive role in my career and for serving on my committee. I appreciate to Dr. Cliff Pereira. I learned a lot from him how to apply theories in class to real world. I wish to thank to Dr. Wilson Schmisseur for serving on my committee.

I especially thank to Dr. Scott Urquhart, who left OSU before I finished, for his encouragement and suggestions. I would like to thank to Dr. Dawm Peters for encouragement and for serving on my committee.

I would like to thank the entire Statistics faculty and staffs for creating a supportive environment. In particular, I wish to thank Maggie, Connie and Beth for patience and conversation.

I could not have completed this degree without the love and faith of my family. A special thanks goes to my wife, Jin, for her love and support. A very special thanks goes to my parents, Bongkwon Park and Sunae Kim, who never lost faith in me.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

GLOSSARY

A^-: Generalized inverse of matrix A

A^+: Moore-Penros inverse of matrix A

ANOVA: Analysis of variance

blue: Best linear unbiased estimator

$\mathcal{F}_d$: Test statistic for unweighted means method

$\mathcal{F}_{\mathrm{R}}$: Test statistic for alternative unweighted means method

gm: Gauss-Markov

K_d: Critical value of unweighted means method

K_{R}: Critical value of alternative unweighted means method

MSD: Usual deviation mean square

MSD_d: Deviation mean square for unweighted means method

$\mathrm{MSD}_{\mathrm{R}}$: Deviation mean square for alternative unweighted means method

MSE: Mean square error

N(A): Null space of matrix A

R(A): Range space of matrix A

P_A: Orthogonal projection operator on range space of matrix A

r(A): rank of matrix A

SSD: Usual deviation sum of squares

SSD_d: Deviation sum of squares for unweighted means method

SSE: Error sum of squares

UANOVA: Unweigted means ANOVA

UM: Unweighted means

TESTING HYPOTHESES USING UNWEIGHTED MEANS

1. INTRODUCTION

In the analysis of linear models for designed experiments, balanced data sets are frequently summarized in an analysis of variance (ANOVA). In balanced data sets the sums of squares in the ANOVA are additive and generally accepted by statisticians as a good way to summarize the data for testing the linear hypotheses that are most often of interest. For unbalanced data there typically does not exist one single unique ANOVA that summarizes the data. The theory of linear models can be applied in fixed linear models to construct an additive ANOVA for testing any linear hypothesis, but this ANOVA changes with each linear hypothesis of interest. Similar statements can be applied to mixed linear models, although the theory for unbalanced mixed linear models does not always lead to a satisfactory answer as it does in the fixed linear model.

This thesis is concerned with a method for forming an ANOVA for unbalanced data sets called the unweighted means ANOVA (UANOVA). The UANOVA procedure was first described by Yates (1934) for an unbalanced two-way model with interaction as an approximate, but computationally simple method of analysis. Today the computationally simple part of a UANOVA justification is not particularly relevant because of high speed computers. Even so, the method is still sometimes recommended in today's literature. It is most popular in mixed linear models as a starting point for computing variance component estimators and confidence intervals. A few of the references in this area are Anderson and Bancroft (1952), Searle (1971), Henderson (1978) and Burdick and Graybill (1992).

It has also been recommended as an alternative way for testing certain linear hypotheses in fixed linear models. The two original references in this direction are Grosslee and Lucas (1965) and Rankin (1974). The theory of using the UANOVA for testing linear hypotheses in fixed effects linear models is less well developed than the UANOVA procedure for variance components in mixed linear models. It is the purpose of this thesis to investigate the UANOVA method for this less well developed area of testing linear hypotheses in fixed effects models.

Much of the literature for testing linear hypotheses in fixed effects models with the UANOVA method is concerned with the two-way model with interaction. This literature is reviewed in Chapter 2. One problem with this literature, which occurs in most of the literature for the two-way model with interaction, is the lack of consensus as to what is the linear hypothesis of no main effect. The recommendations here span the spectrum from those statisticians that say it does not make any sense to test a main effect hypothesis in the model with interaction to something like SAS that provides several alternative sums of squares, e.g., Types I, II and III. It is not the purpose of this thesis to settle this argument, but to ferret out the relevant parts of the literature that concern the UANOVA procedure. For example, some authors have compared the SAS Type II main effect sum of squares (sometimes called the least squares method) with the main effect sum of squares from a UANOVA, but generally we do not believe this is appropriate since the two sums of squares are testing different linear hypotheses. A more relevant comparison is with the Type III sum of squares since the linear hypothesis tested by a main effect Type III sum of squares is the same linear hypothesis as tested by the corresponding main effect sum of squares in a UANOVA.

Although the UANOVA procedure is applicable to unbalanced data sets, there are some limitations. In particular, there cannot be missing cells and continuous variables like covariates cannot be accommodated. In Section 2 of Chapter 4 we deal with this aspect of the UANOVA procedure. We show for a general linear model how the UANOVA procedure can be generalized for any linear hypothesis. This UANOVA generalization, as with the usual linear model test for a linear hypothesis, depends on the linear hypothesis. That is, different linear hypotheses lead to different UANOVAs. Instead of referring to them as a UANOVA we simply refer to the procedure as the UM (unweighted means) method. To accomplish our generalization we require some general facts from linear model theory which are provided in Chapter 3. Our extension of the UM procedure to the general linear model is based on a submodel. This submodel is not unique, but generally it is pretty clear what the submodel should be. Also, in Chapter 4 we investigate the distribution of the test statistic for the UM procedure.

In Chapter 5 we examine several ways for determining the critical value for the UM procedure and provide some power comparisons. In the literature the critical value calculations have been done via a Satterthwaite approximation and the power computations have been done by simulations. However, we were able to adopt the Imhof (1961) methods to do these calculations which essentially provide exact answers, except for any numerical integration and truncation errors.

Traditionally the UM procedure is described for models with interaction, but our generalization is very general and can be used in completely additive models. However, we suspect our UM generalization in an additive model is not what would typically be expected. This is discussed in Section 5 of Chapter 4. We conclude the thesis with some observations in Chapter 6.

2. REVIEW OF THE TWO-WAY MODEL WITH INTERACTION

Consider the two-way classification model with interaction. Let Y_{ijk} be the kth observation in row i and column j for $i = 1, 2, \cdots, a$, $j = 1, 2, \cdots, b$ and $k = 1, 2, \cdots, n_{ij}$. The model, in its unconstrained parametrization form, is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \ , \tag{2.1}$$

where μ is the general mean, α_i is the effect due to row i, β_j is the effect due to column j, γ_{ij} is the interaction effect due to row i and column j and the e_{ijk} are independent random errors with means 0 and variances σ^2 respectively. This feature of having more parameters in the expectation than observed cell means to estimate them from is sometimes called an over-parametrization. However, we shall generally refer to the expectation of Y_{ijk} in model (2.1) as the unconstrained parametrization as opposed to the over-parametrized model. Throughout this chapter we suppose that there are no missing observations (all $n_{ij} > 0$), but that the design can be unbalanced.

In the literature there is a consensus for how to test the linear hypothesis of no interaction in Model (2.1). The devSS (deviation sum of squares) for the null hypothesis of no interaction is the sum of squares for the interaction effects adjusted for all other effects. Let us denote this sum of squares, using the R (reduction) notation, as $R(\gamma|\mu,\alpha,\beta)$. For testing a linear hypothesis of no main effects, however, a similar approach does not work since a quantity like $R(\alpha|\mu,\beta,\gamma)$ is identically zero. As a result, there are several different methods that have been used and/or recommended. The two most popular methods are the "least squares" and "constrained parametrization" methods which are discussed in Sections 2.2 and 2.3 respectively. In the next section the infrequently recommended UM (unweighted means) method is discussed which is the method generalized and investigated in this thesis.

2.1. Method of Unweighted Means.

The UM procedure for model (2.1) was described by Yates (1934) as an approximate, but computationally simple method of analysis. The UM method can also be applied to many other types of experimental designs. A few of the references to this method are Anderson and Bancroft (1952), Gosslee and Lucas (1965), Searle (1971), Rankin (1974), Henderson (1978) and Burdick and Graybill (1992). For model (2.1) let $Y_{dij} = \overline{Y}_{ij.}$. Then we have the model

$$Y_{dij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \overline{e}_{ij.}. \tag{2.2}$$

The UM residual sum of squares is the usual within sum of squares for model (2.1). The remaining UM sums of squares can be computed from model (2.2) acting as though the model is balanced with each Y_{dij} being the average of n_h observations where n_h is the harmonic mean of the actual sample sizes. That is,

$$n_h = \frac{ab}{\sum_i^a \sum_j^b \frac{1}{n_{ij}}}. \tag{2.3}$$

Table 2.1 shows the formulae for the UANOVA (unweighted means ANOVA). Note that generally the only sum of squares in this table that has a χ^2 distribution is the residual sum of squares. As a result, F-ratios formed from this table typically do not follow the F-distribution.

Some writers, e.g., Snedecor and Cochran (1967), have recommended using the UM procedure only when the discrepancies in sample sizes is small, but other writers have not. Gosslee and Lucas (1065) and Rankin (1974) investigated properties of UM in hypothesis testing. Gosslee and Lucas considered model (2.1) and Rankin considered both model (2.1) and the one-way classification model. Rankin concluded that UM is more powerful than the least squares (see next section) when a small sample is taken from the population with the largest mean. Conversely, when the small sample is taken from another population the least squares test is more powerful. Levy, Naura, and Abrami (1975) recommend the use of the UM in the 2^2 experiment over the least squares method. They argued that, although UM is

Table 2.1 UANOVA table for model (2.1)

Source	df	Sum of Squares
A	$a-1$	$\mathrm{SSD_{AU}} = n_h b \sum_{i=1}^{a} (\overline{Y}_{di.} - \overline{Y}_{d..})^2$
B	$b-1$	$\mathrm{SSD_{BU}} = n_h a \sum_{j=1}^{b} (\overline{Y}_{d.j} - \overline{Y}_{d..})^2$
AB	$(a-1)(b-1)$	$\mathrm{SSD_{ABU}} = n_h \sum_{i=1}^{a}\sum_{j=1}^{b} (Y_{dij} - \overline{Y}_{di.} - \overline{Y}_{d.j} + \overline{Y}_{d..})^2$
Residual	$n-ab$	$\mathrm{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} (Y_{ijk} - \overline{Y}_{dij})^2$

approximate, their simulations demonstrate that the approximation is quite good and since UM is simpler, it should be used in preference to least squares. For the Levy, Naura, and Abrami (1975) recommendation, Speed and Monlezun (1979) made the following two arguments. First, for the situation Levy, Naura, and Abrami considered, UM is exact and not approximate. That is, both UM and least squares provide exact tests. Second, the preferred method depends on the hypothesis of interest. Speed, Hocking and Hackney (1978) argued that UM, in general, provides sums of squares that are only approximately chi-squared so that the mean squares in a UANOVA do not correspond to testing linear hypotheses.

The SS from a UANOVA have also been used to construct confidence intervals for variance components. Thomas and Hultquist (1978) discussed how this can be done in the unbalanced random one-way model. They recommended using the SS from a UANOVA as though they came from a balanced design. That is, any method developed for a balanced design could be used by simply substituting the appropriate SS from a UANOVA. This idea was followed up by a number of writers and culminated in a summarization by Burdick and Graybill (1992). Some more recent work along this line can be found Eubank, Seely and Lee (2001) and Purdy (1998). Additionally the UANOVA has been used for estimation of variance components as discussed in Henderson (1978).

2.2. Method of Least Squares (SAS Type II).

The term "least squares method" as used in the previous section refers to a specific way of forming the SS in an ANOVA table for model (2.1). In particular, the main effect sum of squares for rows (factor A) is the sum of squares for the α_i adjusted for the β_j and the sum of squares for columns (factor B) is the sum of squares for the β_j adjusted for the α_i Let us denote these sums of squares using the R notation as $R(\alpha|\mu, \beta)$ and $R(\beta|\mu,\alpha)$ respectively. Then the least squares ANOVA table can be formed as in Table 2.2.

Note that the residual sum of squares is the same in both Tables 2.1 and 2.2, but typically the other sums of squares in the two tables and not equal. Also, note that the sums of squares in the least squares table are equivalent to the Type II SS for model (2.1) generated by SAS GLM. Thus, they are easily calculated.

Table 2.2 ANOVA Table for the Least Squares Method

Source	df	Sum of Squares
A	$a-1$	$R(\alpha\|\mu,\beta)$
B	$b-1$	$R(\beta\|\mu,\alpha)$
AB	$(a-1)(b-1)$	$R(\gamma\|\mu,\alpha,\beta)$
Residual	$n-ab$	SSE

Example 2.1. To illustrate the UM and least squares procedures, consider model (2.1) with $i = 1, 2$ and $j = 1, 2$. Suppose the data are:

Data (Speed and Monlezun (1979))

		B	
	Level	1	2
A	1	10.0 9.8 10.3	25.0 25.1 25.2 24.8 24.6 25.0 25.3 25.1
	2	15.2 15.0 15.0 14.9	19.8 19.9 20.0 20.0 20.4 20.1 19.7

For this data, the ANOVA tables for the two procedures are given in Table 2.3. For the row effects, it is clear from the table that UM fails to reject the null hypothesis, while least squares rejects null hypothesis.

Because of this discrepancy in the conclusions reached by the two procedures, it seems natural to ask if these two methods test same hypotheses. Table 2.4 gives two different row effect null hypotheses and a single null hypothesis for the

Table 2.3 ANOVA Table

Source	df	SS	MS	F	p-value
UM					
Row	1	.0014	.0014	0.03	0.8461
Column	1	467.1091	467.1091	10110.59	
Interaction	1	117.9165	117.9165	2552.65	
Residual	18	.8315	.0462		
least squares					
Row	1	19.1339	19.1339	414.21	<0.0001
Column	1	434.2702	434.2702	9401.05	
Interaction	1	117.9165	117.9265	2552.65	
Residual	18	.8315	.0462		

interaction effect. Note that the column effect hypotheses are omitted in Table 2.4 since they can be easily inferred from the row effect hypotheses.

Table 2.4 Hypotheses to be tested in the 2×2 model

Row Effects

HAU : $\alpha_1 + \overline{\gamma}_{1.} = \alpha_2 + \overline{\gamma}_{2.}$

HA2 : $\frac{n_{11}n_{21}}{n_{11}+n_{21}}(\alpha_1 - \alpha_2 + \gamma_{11} - \gamma_{21}) = \frac{n_{12}n_{22}}{n_{12}+n_{22}}(\alpha_2 - \alpha_1 + \gamma_{22} - \gamma_{12})$

Interaction Effect

HI3 : $\gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}$

By examining the SS in the UM and least squares procedures, it can be determined that the null hypotheses being tested are those given in the following table:

	UM	least squares
Row Effect	HAU	HA2
Interaction Effect	HI3	HI3

Notice that both procedures test the same null hypothesis for the interaction, but that the null hypotheses for the row effects (column effects) can be quite different. Thus, from this table, it is clear that the decision to use one method over the other for testing row effects should be based on the hypothesis of interest. For this particular data set we have

HAU: $\alpha_1 + \overline{\gamma}_{1.} = \alpha_2 + \overline{\gamma}_{2.}$ and

HA2: $\frac{12}{7}(\alpha_1 - \alpha_2 + \gamma_{11} - \gamma_{21}) = \frac{56}{15}(\alpha_2 - \alpha_1 + \gamma_{22} - \gamma_{12})$.

The hypotheses tested by UM are the same hypotheses that are tested in the balanced case. Note that least squares hypotheses are functions of the sample sizes, but that the UM hypotheses are not. □

Speed and Monlezun (1979) emphasized that existing software can be used to obtain a UANOVA. They mentioned that they have converted, with very little effort, BMD02V to handle the a general 2^k experiment with unequal numbers per cell. However, in a 2^k experiment with unequal numbers per cell the UANOVA can be obtained directly from existing software since the Type III ANOVA generated by SAS GLM is the same as the UANOVA in 2^kexperiments.

2.3. The Constrained Parametrization (SAS Type III).

The previous section discussed one of two frequently recommended methods for getting main effect sums of squares for model (2.1). The other can be described in several ways, one of which is through adjusted sums of squares with the usual constrained parametrization. Consider the standard constrained parametrization for model (2.1),

$$\mathrm{E}(Y_{ijk}) = \mu^c + \alpha_i^c + \beta_j^c + \gamma_{ij}^c\,, \text{ all } i\,, j\,, k\,, \tag{2.4}$$

where the parameters are constrained according to the following conditions

$$\Sigma_i\alpha_i^c = 0\ ,\ \Sigma_j\beta_j^c = 0\,,\ \ \Sigma_i\gamma_{ij}^c = 0\ ,\text{ all } i, \text{ and } \ \Sigma_j\gamma_{ij}^c = 0\ ,\text{ all } j\ .$$

Let $R(\alpha^c|\mu^c, \beta^c, \gamma^c)$ be the sum of squares for the α_i^c adjusted for all other parameters. Here the constraints must be taken into account when computing the adjusted sum of squares; otherwise one would compute $R(\alpha|\mu, \beta, \gamma)$ which is identically zero. This sum of squares is the main effect sum of squares for factor A. Similarly, $R(\beta^c|\mu^c, \alpha^c, \gamma^c)$ is the main effect sum of squares for factor B and $\mathrm{R}(\gamma^c|\mu^c, \alpha^c, \beta^c)$is the interaction sum of squares. It can be shown that these adjusted sums of squares for the constrained parametrization (2.4) are identical with the Type III sums of squares from SAS GLM when there are no missing cells in model (2.1). Thus, these sums of squares are easily calculated.

It is easy to check that $R(\gamma^c|\mu^c, \alpha^c, \beta^c) = R(\gamma|\mu, \alpha, \beta)$ so that the least squares method and the constrained parametrization method are testing the same null hypothesis of no interaction. Further, by construction it follows that $\mathrm{R}(\alpha^c|\mu^c, \beta^c, \gamma^c)$ is the devSS for testing the null hypothesis

$$\text{HA3: } \alpha_1^c = \ \dots = \alpha_a^c\ .$$

Recall for the 2^k experiment that a SAS Type III ANOVA is the same as the UANOVA. Thus, for Example 2.1 the UANOVA in Table 2.3 is also the Type III ANOVA and hence HAU in Table 2.4 is the null hypothesis for being tested by $R(\alpha^c|\mu^c, \beta^c, \gamma^c)$.

2.4. The Cell Mean Parametrization.

In the previous sections we have introduced the most frequently recommended (Types II and III) ANOVAs for the two-way model with interaction as well as the UANOVA. For the 2 × 2 model, we have also seen through Example 2.1 what hypotheses are being tested in each of these ANOVAs. To examine what is being tested in the general model, it is convenient to introduce the cell mean parametrization used by some writers such as Kutner (1974) and Speed, Hocking and Hackney (1978). In particular, the cell mean parametrization is given by

$$\mathrm{E}(Y_{ijk}) = \mu_{ij} \ , \ \text{all} \ i, \ j, \ k \ , \tag{2.5}$$

where μ_{ij} is defined as the population mean of cell $i, \ j$.

By expressing the previous null hypotheses in terms of the cell mean parametrization, it is easy to distinguish the differences among the various hypotheses. For example, by using the relationships

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \ = \mu^c + \alpha_i^c + \beta_j^c + \gamma_{ij}^c \ , \tag{2.6}$$

one can express HA2 and HA3 for a 2^2 experiment as

HA2: $\frac{n_{11}n_{21}}{n_{11}+n_{21}}(\mu_{11} - \mu_{21}) = \frac{n_{12}n_{22}}{n_{12}+n_{22}}(\mu_{22} - \mu_{12})$ and

HA3: $\overline{\mu}_{1.} = \overline{\mu}_{2.}$.

For the general model, the various null hypotheses are contained in Table 2.5. The hypothesis HA1 in this table is associated with $R(\alpha|\mu)$ which can be obtained as a Type I SS in SAS. The hypotheses HA2 and HA3 are associated with SAS type II and III sums of squares and HAU is the null hypothesis for no row effects in a UANOVA. For completeness, column effect hypotheses are also included in the table and have a similar interpretation as the row effect hypotheses. The interaction sum of squares is the same in Types I, II and III sums of squares and the associated null hypothesis is HI3. In the UANOVA, the null hypothesis HIU is the same as the HI3 null hypothesis, but in general the two sums of squares are not the same. Justification for the Type I, II and III null hypothesis expressions can be found in Speed, Hocking and Hackney (1978, JASA) while justification for the UANOVA null hypothesis expressions is easily inferred later from Example 4.4

Table 2.5 Hypotheses in the two-way classification model

Row Effects	HA1: $\sum_j n_{ij}\mu_{ij}/n_{i.} = \sum_j n_{i'j}\mu_{i'j}/n_{i'.}$
	HA2: $\sum_j n_{ij}\mu_{ij} = \sum_{i'}\sum_j \frac{n_{ij}n_{i'j}\mu_{ij}}{n_{.j}}$
	HA3, HAU: $\overline{\mu}_{i.} = \overline{\mu}_{i'.}$
Column Effects	HC1: $\sum_i n_{ij}\mu_{ij}/n_{.j} = \sum_i n_{ij'}\mu_{ij'}/n_{,j'}$
	HC2: $\sum_i n_{ij}\mu_{ij} = \sum_{j'}\sum_i \frac{n_{ij}n_{ij'}\mu_{ij'}}{n_{i.j}}$
	HC3, HCU: $\overline{\mu}_{.j} = \overline{\mu}_{.j'}$
Interaction Effect	HI3, HIU: $\mu_{ij} - \mu_{ij'} - \mu_{i'j} + \mu_{i'j'}$

The following table summarizes the previous discussion about which hypotheses are tested by each of the methods. The table does not contain the column effect information, but this information is easily determined from the row effect information.

	Row	Row SS	Interaction
Type I	HA1	$R(\alpha\mid\mu)$	HI3
Type II	HA2	$R(\alpha\mid\mu,\beta)$	HI3
Type III	HA3	$R(\alpha^c\mid\mu^c,\beta^c,\gamma^c)$	HI3
UM	HA3	SSAU	HI3

A few observations about Table 2.5 seem appropriate. First, the table assumes that all of the n_{ij} are nonzero. Second, when the design is balanced, i.e., all of the n_{ij} are equal, then the entries in each column reduce to the same quantity. However, when the n_{ij} are unequal the row effect sums of squares are generally all different and the two interaction sums of squares are also typically different. Third, notice that Types I and II row effect null hypotheses depend on the cell sizes whereas HA3 and HAU are independent of the cell sizes. Fourth, with respect to the Type II row effect SS, it should be mentioned that frequently when this hypothesis is recommended, it is only recommended when the interaction SS is not significant, e.g., see Searle (1971).

Note that some authors refer to the Type II method as "fitting constants" or "least squares." Such a title is misleading since all of the methods, except that of unweighted means, are based on least squares fits of some model. (Speed, Hocking and Hackney, 1978).

Previously we have outlined four possible methods for obtaining a row effect sum of squares for model (2.1). The preferred method should depend on the hypothesis of interest. Francis (1973) stated that HA3 is the type of hypothesis that most people really want to do and Kutner (1974) also stated that HA3 is generally the more appropriate hypothesis. Several authors compared UM with the Type II method. For example, Levy, Laura and Abrami (1975) performed simulations in 2^2 experiments for such comparisons. Rankin (1974) compared UM with the Type II method in the one-way classification model. In the one-way classification model, Types I, II and III all give the same sum of squares for the linear hypothesis of no treatment effects. However, in other models it should be noted that such comparisons are somewhat suspect since different linear hypotheses are being tested.

Speed, Charles and Monlezun (1979) argue that for 2^2 experiments UM is testing a reasonable hypothesis, but Type II is generally not. Note that the associated sum of squares from UM and Type III are testing the same hypotheses, which is the same hypothesis tested in the balanced case In fact, in 2^2 experiments, the associated sum of squares from UM and Type III are identical. Therefore, UM and Type III can be compared, however their comparison has not been explored. Note that although UM and Type III methods test the same null hypotheses, they generally have different power functions.

3. REVIEW OF LINEAR MODELS

This chapter summarizes the elements of linear model theory that are deemed important to understanding the ideas in the succeeding pages. Suppose that Y is an $n \times 1$ random vector. In addition to the random vector Y, the components of a linear model generally consist of four items: An assumption about the expectation vectors $\mathrm{E}(Y)$, an assumption about the covariance structure $\mathrm{Cov}(Y)$, an assumption about any functional relationships between the mean vector and the covariance structure and an assumption about the distribution of Y. In this thesis we will always suppose that there are no functional relationships between the mean vector and the covariance structure. Also, the distributional assumptions for the first few sections will be simply first and second moment assumptions, but when we discuss the residual and deviation sum of squares we will assume that Y has a multivariate normal distribution. Note that these assumptions are not limiting assumptions in the sense that they are the usual assumptions made in linear models.

3.1. Expectation, Estimability and Parametrizations.

The set consisting of all possible expectation vectors $\mathrm{E}(Y)$ is called the *expectation space* and is denoted by Ω. For our purposes, we suppose that Ω is a subspace. While this is not necessary, it does make notation more convenient and allows us to circumvent a couple of irregularities that arise when Ω is allowed to be an arbitrary set. A *parametrization* for E(Y) is an expression of the form

$$\mathrm{E}(Y) = X\delta,\ \Delta'\delta = 0, \tag{3.1}$$

where δ is a $p \times 1$ vector of unknown parameters, X is a known $\mathrm{n} \times \mathrm{p}$ matrix, Δ is a known $p \times s$ matrix and $\Omega = \{X\delta : \Delta'\delta = 0\}$. Note that in theory, one starts with the expectation space and then comes the parametrization. In practice, however, one starts with a given parametrization which then defines the expectation space from which other parametrizations can be defined. The matrix X in a parametrization is called the *design matrix* (although some authors restrict design matrix to mean that X is composed of zero's and one's) and N(Δ'), the null space of Δ', is called the *parameter space.* Note that the dimension of Ω can be expressed as

$$m = \dim\Omega = \mathrm{r}(X', \Delta) - \mathrm{r}(\Delta)\,,$$

where $\mathrm{r}(A)$ denotes the rank of a matrix A.

For the most part, parametrizations considered in the sequel are unrestricted. This is usually indicated by a statement such as "δ unknown" which means that $\mathbb{R}^p$ is the parameter space so that $\Delta = 0$. In this situation $\Omega = \mathrm{R}(X)$ where $\mathrm{R}(A)$ denotes the range or column space of the matrix A and $m = \mathrm{r}(X)$.

Example 3.1 (two-way model with interaction). Let Y denote the $n = \Sigma_{ij} n_{ij} \times 1$ vector composed of the Y_{ijk} in model (2.1). Write the expectation in matrix form as

$$\mathrm{E}(Y) = 1\mu + A\alpha + B\beta + T\gamma = X\delta,\ \delta \text{ unknown.}$$

Here $p = 1 + a + b + ab$ and since there are no restrictions we have $\Delta = 0$ so that $\Omega = \mathrm{R}(X)$ and $m = \mathrm{r}(X) = ab$. Now let Δ' be such that $\Delta'\delta^c = 0$ describes the constraints in the parametrization (2.4). For example, if $a = b = 2$, then

$$\Delta' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Then it can be shown that $\{X\delta^c : \Delta'\delta^c = 0\}$ is equal to R(X) so that (2.4) is also a parametrization. Finally, it is easy to check that R(X) = R(T) so that the unrestricted cell means in (2.5) also constitute a parametrization. Although the n_{ij} were all assumed to be nonzero for model (2.4), it should be noted here that the parametrizations in this example do not require this assumption. However, the value of $m = ab$ does require nonzero n_{ij}. In general m is equal to the number of nonzero n_{ij}.□

Example 3.2 (two-way additive model). Consider the two-way classification model without interaction. This is the same setup as in model (2.1) except that the γ_{ij} interaction terms do not appear in the model. In particular, the unconstrained parametrization for this model is

$$\mathrm{E}(Y_{ijk}) = \mu + \alpha_i + \beta_j \,,$$

where $i = 1,\ldots,a$, $j = 1,\ldots b$ and $k = 1,\ldots,n_{ij}$. And the matrix version is

$$\mathrm{E}(Y) = 1\mu + A\alpha + B\beta = X\delta,\ \delta \text{ unknown},$$

where $X = (1, A, B)$. As in Chapter 2 we assume all of the n_{ij} are nonzero, but this is not necessary for the general model. The constrained parametrization and the cell mean parametrization are also defined for this model. The constrained parametrization is like parametrization (2.4) without the interaction terms. To illustrate in matrix form, suppose that $a = 2$ and $b = 2$. Then the constrained parametrization is

$$\mathrm{E}(Y) = X\delta^c, \Delta'\delta^c = 0,$$

where

$$\Delta' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

To get the cell mean parametrization for the additive model one must add restrictions to the parametrization (2.5). To illustrate, suppose again that $a = b = 2$. Then the cell means parametrization can be written as

$$\mathrm{E}(Y_{ij}) = \mu_{ij} \quad ,$$

where $$\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} \ = 0.$$

In matrix form the parametrization can be written as

$$\mathrm{E}(Y) = T\mu, \Gamma'\mu = 0,$$

where

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mu = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \text{ and } \Gamma' = [\,1 \quad -1 \quad -1 \quad 1\,].$$

Note that all of the expressions above are parametrizations for $\mathrm{E}(Y)$ because it can be shown that $\Omega = \mathrm{R}(X) = \{X\delta^c : \Delta'\delta^c = 0\} = \{T\mu : \Gamma'\mu = 0\}$.□

One part of the theory of linear models is the study of linear estimation. That is, the study of linear estimators to estimate linear parametric vectors. By a *linear estimator* we mean a vector $\mathrm{A'Y}$ where A is a known matrix and by a *linear parametric vector* we mean a vector of the form $\Lambda'\delta$ where Λ is a known matrix. Since our attention in the sequel will be exclusively on linear estimators and linear parametric vectors, the term "linear" is frequently omitted.

Definition. A parametric vector $\Lambda'\delta$ is said to be *estimable* provided that there exists a linear estimator that is unbiased for $\Lambda'\delta$.

The following well known proposition characterizes estimability.

Proposition 3.1.

(a) A parametric vector $\Lambda'\delta$ is estimable if and only if $\mathrm{R}(\Lambda) \subset \mathrm{R}(X',\Delta)$.

(b) The parametric vector δ is estimable if and only if $\mathrm{r}(X', \Delta) = p$.

In the previous interaction example, the parametric vector in the unconstrained parametrization is not estimable whereas it is estimable in both the constrained parametrization and the cell mean parametrization whenever the n_{ij} are all nonzero. If, however, some of the n_{ij} are zero, then the parametric vector in each of the three parametrizations is not estimable. In the additive example, the situation is somewhat changed. In particular, the parametric vector in the unconstrained parametrization is not estimable, but the parametric vectors in the constrained parametrization and the cell mean parametrization is estimable whenever $\mathrm{r}(X) = a + \mathrm{b} - 1$ which is the same as saying the design matrix X is connected.

3.2. Covariance Structure, *blues,* and *gm* Vectors.

The previous section discussed the first moment linear model assumptions and the notion of estimability. The present section is devoted to the second moment assumptions and linear estimation. Throughout this section suppose that

$$\mathrm{E}(Y) = X\delta\,,\ \Delta'\delta = 0,$$

is a parametrization for $\mathrm{E}(Y)$.

Definitions.

- ***blue*** (best linear unbiased estimator). A linear estimator $\mathrm{T}'Y$ is said to be a *blue* (for its expectation) provided that

  $$\mathrm{Cov}(T'Y) \leq \mathrm{Cov}(A'Y)$$

  for all $\mathrm{Cov}(Y)$ and for all linear estimators $A'Y$ that have the same expectation as $T'Y$. It is said to be a *blue* for $\Lambda'\delta$ provided that it is a *blue* and it is unbiased for $\Lambda'\delta$.

- ***gm*** (Gauss-Markov) **for δ.** A random vector $\widehat{\delta}$ that is a function of Y is said to be *gm* for δ provided that $\Lambda'\widehat{\delta}$ is a *blue* for $\Lambda'\delta$ whenever $\Lambda'\delta$ is estimable.
- Y is said to be proper provided that it's sample space is a subset of the range of its covariance matrix.

The notions of a *blue* and *gm* for δ are standard linear model definitions, but the notion of Y being proper is not. Note that if $\mathrm{Cov}(Y)$ is positive definite, then its range is $\mathbb{R}^n$ so that Y is automatically proper. Thus, the notion of proper only comes into play when the covariance matrix in a linear model is singular. All of the basic models we consider will have a positive definite covariance matrix so they will be proper. If fact, the basic linear models that we will consider will have $\mathrm{Cov}(Y) = \sigma^2 I$. However, even in this context it is easy to come up with another linear model that does have a singular covariance matrix as illustrated in the following example.

Example 3.3 (partitioned model). Suppose that δ is unknown, that $X\delta$ is partitioned as $X\delta = X_1\delta_1 + X_2\delta_2$ and that $\mathrm{Cov}(Y) = \sigma^2 I$. If one is only interested in estimable parametric vectors involving δ_2, then it is natural to consider the linear model $Z = N_1 Y$ where N_1 is the orthogonal projection operator on $\mathrm{N}(X_1')$. Notice that

$$\mathrm{E}(Z) = N_1 X_2 \delta_2 \quad \text{and} \quad \mathrm{Cov}(Z) = \sigma^2 N_1 \ .$$

That is, the Z linear model has a singular covariance matrix. However, notice that the sample space of Z is $\{N_1 y : y \text{ in sample space of } Y\}$ which is contained in $\mathrm{R}(N_1)$ so that the Z linear model is proper.□

The above example illustrates that singular covariance matrices can easily arise in standard linear models and thus it would be helpful to know how to deal with them. There is a general theory of linear models for singular covariance matrices, but that theory is a bit cumbersome. However, if the linear model is proper, then the basic theory for a positive definite covariance structure is essentially unchanged as illustrated by the following proposition.

Proposition 3.2. Suppose that δ is unknown, that $\mathrm{Cov}(Y) = \sigma^2 V$ and that Y is proper. Let G be a g-inverse for V. Then the following statements can be made.

(a) $T'Y$ is a *blue* if and only if $\mathrm{R}(VT) \subset \Omega$.

(b) A random vector $\widehat{\delta}$ is *gm* for δ if and only if $\mathrm{X}'GX\widehat{\delta} = \mathrm{X}'GY$.

(c) If $\Lambda'\delta$ and $\Gamma'\delta$ are estimable and if $\widehat{\delta}$ is *gm* for δ, then

$$\mathrm{Cov}(\Lambda'\widehat{\delta}, \Gamma'\widehat{\delta}) = \sigma^2\Lambda'(\mathrm{X}'GX)^{+}\Gamma\,.$$

The facts in the above proposition are well known when V is positive definite. In fact, (a) is known to be true whether or not Y is proper. Also, (c) is known to be true whenever $\Omega \subset \mathrm{R}(V)$ (which is a consequence of proper). This condition for (c) also implies that $\widehat{\delta}$ in (b) is *gm,* but the converse in (b) does need proper to insure validity. Recall in the above partitioned model example that the Z linear model was proper. Also, note that $N_1$is an orthogonal projection operator so that it is a g-inverse for itself. Thus, from (b) we can say that $\widehat{\delta}_2$ is *gm* for δ_2 with respect to the Z linear model if and only if $\widehat{\delta}_2$ satisfies the reduced normal equations $X'N_1X\widehat{\delta}_2 = X'N_1Y$. In the partitioned model example we demonstrated directly that Z was proper. However, we could conclude this fact immediately from the following lemma.

Lemma 3.3. If Y is proper, then the linear model for $Z = A'Y$ is proper where A is any given known matrix.

This lemma, combined with the previous proposition is very useful in linear models. The lemma says that any linear transformation of a proper random vector is proper. Hence, since all of our basic linear models will have a positive definite covariance matrix, it follows that any linear model we consider that is derived as a linear transformation of Y will be proper. Also, notice that the above proposition is for the situation when there are no constraints on the parameter vector, i.e., $\Delta = 0$. There are a number of similar results for arbitrary Δ, but we mention only one of them.

Corollary 3.4. Suppose that $\text{Cov}(Y) = \sigma^2 V$ and that Y is proper. Set $U = XB$ where B is such that $\text{R}(B) = \text{N}(\Delta')$. Then $\text{E}(Y) = U\theta$, θ unknown, is a parametrization for $\text{E}(Y)$. Furthermore,

(a) If $\widehat{\theta}$ is *gm* for θ, then $\widehat{\delta} = B\widehat{\theta}$ is *gm* for δ.

(b) If $\Lambda'\delta$ and $\Gamma'\delta$ are estimable and if and if $\widehat{\delta}$ is gm for δ, then

$$\text{Cov}(\Lambda'\widehat{\delta}, \Gamma'\widehat{\delta}) = \sigma^2 \Lambda' B(U'V^{+}U)^{+} B'\Gamma.$$

Notice that to use this corollary you first apply the previous proposition to the parametrization $\text{E}(Y) = U\theta$, θ unknown. That is, $\widehat{\theta}$ and $(U'V^{+}U)^{+}$ would be outputs from the proposition.

This concludes our brief discussion on linear estimation. In summary, we note that in general our basic model will typically have $\text{Cov}(Y) = \sigma^2 I$ so that our basic model and any derived models that are a linear transformation of our basic model will be proper. Also, we will typically have no constraints on our model so that the *gm* equations reduce to the familiar normal or least squares equations $X'X\widehat{\delta} = X'Y$.

3.3. Linear Hypotheses.

Suppose that Y is such that $\mathrm{E}(Y) \in \Omega$ where Ω is a subspace. Then a linear hypothesis is any statement that can be reduced to the form

$$\text{H: } \mathrm{E}(Y) \in \Omega_H \ [\ \Omega_H \text{ is a subspace of } \Omega\ ] \tag{3.2}$$

$$\text{A: } \mathrm{E}(Y) \in \Omega_A, [\ \Omega_A = \{\mu\text{: } \mu \in \Omega\ ,\ \mu \notin \Omega_{\mathrm{H}}\ ].$$

In this statement it is assumed that the covariance structure of Y under the null and alternative hypotheses is the same as in the model for Y. Thus, a *linear hypothesis* is simply a null hypothesis that specifies a linear model whose expectation space is a subspace of Ω. One way to form a linear hypothesis is the following:

$$\text{H: } \mathrm{E}(Y) = X_H\delta_H,\ \Delta_H'\delta_H = 0\,. \tag{3.3}$$

By such a statement we mean (3.2) with $\Omega_H = \{X_H\delta_H\ :\ \Delta_H'\delta_H = 0\}$. We refer to (3.3) as a linear hypothesis on $\mathrm{E}(Y)$ provided that $\Omega_H \subset \Omega$.

Our primary interest is in forming linear hypotheses via a statement about the parameters in a parametrization for $\mathrm{E}(Y)$. Suppose that $\mathrm{E}(Y) = X\delta$, $\Delta'\delta = 0$, is a parametrization for $\mathrm{E}(Y)$. Then our interest will be in forming a linear hypothesis via a statement like

$$\text{H: } \Lambda\,'\delta = 0\ \ (\Delta'\delta = 0) \tag{3.4}$$

$$\text{A: } \Lambda'\delta \neq 0\ \ (\Delta'\delta = 0)\ .$$

In order for this to be a linear hypothesis as in (3.2) we first need $\Omega_H = \{X\delta : \Lambda'\delta = 0, \Delta'\delta = 0\}$ to be a subspace of Ω. This is clearly true so that H: $\mathrm{E}(Y) \in \Omega_H$ is a linear hypothesis on $\mathrm{E}(Y)$. The alternative to this linear hypothesis is A: $\mathrm{E}(Y) \in \Omega_A = \{\mu\text{: } \mu \in \Omega\ ,\ \mu \notin \Omega_{\mathrm{H}}\}$, but we would like the alternative to be specified as in (3.4). That is, we want to interpret the alternative in terms of the parameter vector δ instead of $\mathrm{E}(Y)$. In particular, this means that we need $\Omega_A = \Omega_{\mathrm{A}}^*$ where $\Omega_{\mathrm{A}}^* = \{X\delta : \Lambda'\delta \neq 0\,, \Delta'\delta = 0\}$. In order for this to be true it can be shown that we need $\Lambda'\delta$ to be estimable. We refer to (3.4) as a linear hypothesis on δ provided that $\Lambda'\delta$ is estimable. Typically we shorten a linear hypothesis on δ to a statement like H: $\Lambda'\delta = 0$, but the actual statement is that described in (3.4).

Above two different methods for describing a linear hypothesis have been discussed. Neither of these methods is more general than the other in the sense that any linear hypothesis (3.2) can be described as a linear hypothesis on E(Y) and it can also be described as a linear hypothesis on δ. Thus, any linear hypothesis on δ can be expressed as a linear hypothesis on E(Y) and conversely. Expressing a linear hypothesis on δ as a linear hypothesis on E(Y) is straightforward. For example, take $X_H = X$ and $\Delta_H = (\Lambda,\Delta)$. The converse of expressing a linear hypothesis on E(Y) as a linear hypothesis on δ can always be accomplished, but that requires more work. This is illustrated in the following example.

Example 3.4 (Two-way additive unconstrained model). Consider the two-way additive model with $a = b = 2$ and with $n_{ij} = 1$ for all i, j. Write the model in matrix form using the unconstrained parametrization as

$$E(Y) = 1\mu + A\alpha + B\beta = X\delta.$$

Then H: $\mathrm{E}(Y_{ij}) = \mu_H + \beta_{Hj}$ is a linear hypothesis on $\mathrm{E}(Y)$ because $\Omega_H = \mathrm{R}(1,B)$ is a subspace of $\Omega = \mathrm{R}(X)$. Let us express this linear hypothesis on E(Y) as a linear hypothesis on δ. That is, find an estimable parametric vector $\Lambda'\delta$ such that $\Omega_H = \{X\delta \,:\, \Lambda'\delta = 0\}$. Typically the easiest way to do this is via ad hoc type methods, but a prescribed method that will always work is the following. Find Q such that $\mathrm{N}(Q') = \Omega_H$. Then $\theta \in \Omega_H \Leftrightarrow Q'\theta = 0$. However, it is also true that $\Omega_H \subset \Omega$ so that $\theta \in \Omega_H$ must be of the form $\theta = X\delta$ for some δ. Therefore, we can write $\Omega_H = \{X\delta \,:\, Q'X\delta = 0\}$. Let $\Lambda' = Q'X$.

Since $\Lambda'\delta$ is estimable because $\underline{R}(\Lambda) \subset \underline{R}(X'Q) \subset \underline{R}(X')$, it follows that H:$\Lambda'\delta = 0$ is a linear hypothesis on δ that is equivalent to H: E$(Y) \in \Omega_H$. To illustrate for our particular example, $Q = I - P_B$ is a satisfactory choice since $\Omega_H =$ R(B). This gives

$$Q = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad \Lambda' = Q'X = \begin{bmatrix} 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Therefore hypothesis on E(Y), H: $E(Y_{ij}) = \mu_H + \beta_{Hj}$, is same the hypothesis H: $\alpha_1 - \alpha_2 = 0$ which is a hypothesis on δ.□

3.4. The Residual and Deviation Sum of Squares.

Suppose in this section that E$(Y) = X\delta$, $\Delta'\delta = 0$, is a parametrization for E(Y) and that Cov$(Y) = \sigma^2 V$ where V is a positive definite matrix. Since Y is proper, note that any linear transformation of Y is also proper. The purpose of this section is to briefly review the usual test statistic for testing a linear hypothesis.

The first quantity we need is the residual or error sum of squares. In particular,

$$\text{SSE} = \min_{\mu\in\Omega} ||Y - \mu||_C^2 \tag{3.5}$$

where $||x||_C = x'V^{-1}x$ is called the residual or error sum of squares. Next we need the following proposition. The notation χ^2(k,λ) denotes the noncentral chi-squared distribution with k degrees of freedom and noncentrality parameter λ.

Proposition 3.5. Suppose that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ. Let $\widehat{\delta}$ be *gm* for δ, let D be such that Cov$(\Lambda'\widehat{\delta}) = \sigma^2$D and let $r =$ r(D). Set

$$\text{SSD} = (\Lambda'\widehat{\delta})'\text{D}^-(\Lambda'\widehat{\delta}). \tag{3.6}$$

Recall that $m = \dim\Omega$. If Y has a multivariate normal distribution, then the following statements may be made:

(a) $\mathrm{SSE}/\sigma^2 \sim \chi^2(f)$ where $f = n - m$.

(b) $r = \mathrm{r}(\Lambda,\Delta) - \mathrm{r}(\Delta)$.

(c) $\mathrm{SSD}/\sigma^2 \sim \chi^2(r, \lambda)$ where $\lambda = (\Lambda'\delta)'\mathrm{D}^-(\Lambda'\delta)/\sigma^2$.

(d) The noncentrality parameter λ in (c) is zero if and only if $\Lambda'\delta = 0$.

(e) SSE and SSD are independent.

(f) If Q is a continuous function of Y such that $\mathrm{Q}/\sigma^2 \sim \chi^2(r)$ under the null hypothesis and Q is independent of SSE, then Q = SSD.

This proposition provides all of the necessary information for constructing a test statistic for a linear hypothesis on δ. In particular notice that if Y has a multivariate normal distribution, then

$$\mathrm{SSD}/r \div \mathrm{SSE}/f$$

has a central F-distribution with r and f degrees of freedom under the null hypothesis and a non-central F-distribution under the alternative hypothesis. Several other items seem worth mentioning: (i) Notice that (f) implies that SSD does not depend on the choice of the g-inverse. Similarly the non-centrality parameter also does not depend on the choice of the g-inverse. (ii) The linear hypothesis H: $\Lambda'\delta$ can equivalently be expressed as H: $\Lambda_r'\delta = 0$ provided that $\mathrm{R}(\Lambda_r, \Delta) = \mathrm{R}(\Lambda, \Delta)$. In particular, if Λ_r is also chosen to have linearly independent columns and $\mathrm{R}(\Lambda_r) \cap \mathrm{R}(\Delta) = \{0\}$, then D in the proposition is a positive definite matrix. (iii) There is another basic way to describe SSD. In particular, the difference $\mathrm{SSD} = \mathrm{SSE}_H - \mathrm{SSE}$ where SSE_H is the residual sum of squares under the null hypothesis. Similarly, the degrees of freedom can be expressed as $r = f_H - f$ where $f_H = \dim\Omega_H$. These equalities leads to the familiar ANOVA table.

Table 3.1 ANOVA for testing H: $\Lambda'\delta = 0$

Source of Variation	df	SS	E(MS)
Deviation from H	r	SSD	$\sigma^2 + (\Lambda'\delta)'D^-(\Lambda'\delta)/r$
Residual	f	SSE	σ^2
Residual under H	f_H	SSE_H	

4. GENERALIZED UNWEIGHTED MEANS

This chapter generalizes an unweighted means sum of squares and formalizes that procedure. Some properties of the unweighted means procedure are explored. These include consistency of an unweighted means sum of squares ant the distribution of that sum of squares. In general, an unweighted means sum of squares does not follow a chi-squared distribution, but instead a linear combination of independent chi-squared random variables. We also discuss the test statistics in utilizing an unweighted means sum of squares.

4.1. Two-way with Interaction Revisited.

To introduce the sum of squares for the UM procedure it is convenient to first review, in matrix form, the two-way model with interaction of Chapter 2. Let Y denote the vector of Y_{ijk} of model (2.1) and write the model in the unconstrained parametrization form as in Example 3.1. That is,

$$\mathrm{E}(Y) = 1\mu + A\alpha + B\beta + T\gamma = \mathrm{X}\delta,\ \delta \text{ unknown},$$

where $\mathrm{Cov}(Y) = \sigma^2 I$. With this formulation, the cell means can be expressed as $Y_d = (T'T)^{-1}T'Y$ and model (2.2) can be written in matrix form as

$$\mathrm{E}(Y_d) = 1_{\mathrm{d}}\mu + A_{\mathrm{d}}\alpha + B_{\mathrm{d}}\beta + T_{\mathrm{d}}\gamma = X_{\mathrm{d}}\delta,\ \delta \text{ unknown},$$

where $\mathrm{Cov}(Y_d) = \sigma^2 V$ where $V = (T'T)^{-1}$. Note here that X_{d} is actually the matrix composed of the "distinct" rows of X and that $X = TX_{\mathrm{d}}$.

To obtain the devSS (deviation sum of squares) for the UM procedure, suppose that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ and that $\widehat{\delta}$ is *gm* for δ. Note that $\mathrm{Cov}(\Lambda'\widehat{\delta}) = \sigma^2\mathbb{C}$ where $\mathbb{C} = \Lambda'(X'X)^{+}\Lambda$ so that $\mathrm{SSD} = \widehat{\delta}\,'\Lambda\mathbb{C}^{+}\Lambda'\widehat{\delta}$

is the usual devSS for testing H: $\Lambda'\delta = 0$. The devSS for the UM procedure has this same form except that the design matrix for the cell mean model X_d is used in place of X in the $\mathbb{C}$ matrix. That is, set $\mathbb{C}_d = \Lambda'(X_d'X_d)^{+}\Lambda$. Then

$$\mathrm{SSD}_d = \widehat{\delta}\,'\Lambda\mathbb{C}_d^{+}\Lambda'\widehat{\delta}$$

is the devSS for the UM method. Actually, this devSS must be multiplied by the harmonic mean to make the devSS compatible with SSE. However, we will not consider this aspect of the UM method until later.

To illustrate the above, consider a 3×3 experiment. Then the devSS for the row effect in Table 2.1 can be expressed as SSD_d where

$$\Lambda' = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & -1/3 & -1/3 & -1/3 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & -1/3 & -1/3 & -1/3 \end{bmatrix}$$

This sum of squares can also be expressed in terms of other parametrizations. For example, consider the cell mean parametrization of (2.5). In matrix form this reduces to $\mathrm{E}(Y) = T\mu$, μ unknown, or in terms of the cell means as $\mathrm{E}(Y_d) = \mathrm{T}_d\mu$, μ unknown. If H: $\Gamma'\mu = 0$ is a linear hypothesis on μ, then the usual devSS is $\widehat{\mu}'\Gamma\mathbb{C}^{+}\Gamma'\widehat{\mu}$ where $\mathbb{C} = \Gamma'(T'T)^{-1}\Gamma$ and the devSS for the UM procedure would be this same expression with T replaced by T_d. For example, for the devSS for rows in Table 2.1 with the cell mean parametrization one could use

$$\Gamma' = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}$$

Because $T_d = I$, the devSS for the UM method with the cell mean parametrization reduces to $\widehat{\mu}\,'P_\Lambda\widehat{\mu}$, and this expression (apart from the harmonic mean) is easily seen to be equal to the row effect sum of squares in Table 2.1.

4.2 UM Sum of Squares

In the previous section we saw a fairly general way to describe the devSS for the UM procedure. In this section we formalize that procedure. Suppose that we have the linear model structure

$$\mathrm{E}(Y) = X\delta,\ \Delta'\delta = 0,\ \text{ and } \ \mathrm{Cov}(Y) = \sigma^2 I\ .$$

Now suppose that $X = TX_d$ where T is an $n \times d$ matrix with rank d. This is slightly more general than the cell mean model where X_d would consist of the distinct rows of X and T would be a classification matrix, but the results below only require that T have full column rank. For completeness we note that a *classification matrix* is a matrix whose entries are all 0's and 1's with precisely one 1 in each row and at least one 1 in each column.

Now let $Y_d = (T'T)^{-1}T'Y$. Then the linear model for Y_d is given by

$$\mathrm{E}(Y_d) = X_d\delta, \Delta'\delta = 0, \quad \text{and} \quad \mathrm{Cov}(Y_d) = \sigma^2 V$$

where $V = (T'T)^{-1}$. Several relationships between the Y-model and the Y_d-model should be noted.

- Because T has full column rank, it follows that $\mathrm{R}(X') = \mathrm{R}(X_d')$. Among other things, this implies that $\Lambda'\delta$ is estimable in the Y_d-model if and only if it is estimable in the Y-model.
- Because $\Omega \subset \mathrm{R}(T)$, it follows that the Y_d-model is linearly sufficient. In particular, this means that if $\widehat{\delta}$ is *gm* for δ in the Y_d-model, then it is *gm* for δ in the Y-model.

Now suppose that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ, that $\widehat{\delta}$ is *gm* for δ and that B is any matrix satisfying $\Delta'B = 0$ and $\mathrm{r}(XB) = \mathrm{r}(X', \Delta) - \mathrm{r}(\Delta)$. Set

$$\mathbb{C} = \Lambda' B(B'X'XB)^{+}B'\Lambda\,.$$

Then $\mathrm{Cov}(\Lambda'\widehat{\delta}) = \sigma^2\mathbb{C}$ so that $\mathrm{SSD} = \widehat{\delta}\,'\Lambda\mathbb{C}^{+}\Lambda'\widehat{\delta}$ is the usual devSS for testing H: $\Lambda'\delta = 0$. Let $\mathbb{C}_d$ denote the above $\mathbb{C}$ matrix with X replaced by X_d. Then, as in the previous section,

$$\mathrm{SSD}_d = \widehat{\delta}\,'\Lambda\mathbb{C}_d^{+}\Lambda'\widehat{\delta}\ , \quad \mathbb{C}_d = \Lambda' B(B'X_d'X_dB)^{+}B'\Lambda \tag{4.1}$$

is the devSS for the UM procedure. Note in the expressions for $\mathbb{C}$ and $\mathbb{C}_d$, as well as the two devSS expressions, that we have used Moore-Penrose inverses. It is known for $\mathbb{C}$ and SSD that Moore-Penrose inverses can be replaced by any g-inverses and the resulting expressions are the same. This is also true for $\mathbb{C}_d$ and SSD_d as will be established in the next section. Even so, for convenience and to eliminate the perception of non-uniqueness via an arbitrary g-inverse, we will use Moore Penrose inverses.

Notice that SSD_d is defined via the matrix B which is not unique. Hence, a reasonable question is whether or not SSD_d is well defined. The same question can be asked about SSD, but this question is easily answered. In particular, since $\Lambda'\widehat{\delta}$ is unique it follows that its covariance matrix is unique so that $\mathbb{C}$ is unique. That is, any matrix B satisfying the stated conditions leads to the same $\mathbb{C}$. Basically the same argument can be used for SSD_d. That is, momentarily suppose that Y_d has a covariance matrix of the form $\sigma^2 I$. Then $\mathbb{C}_d$ would be the covariance matrix of $\Lambda'\widehat{\delta}$ and hence is uniquely defined no matter what B is selected.

With respect to the matrix B used in defining $\mathbb{C}$ and $\mathbb{C}_d$, the following linear model facts are sometimes handy.

- A satisfactory choice for B is any matrix satisfying $\mathrm{R}(B) = \mathrm{N}(\Delta')$. For example, $B = I - P_\Delta$ satisfies the conditions or if $\Delta = 0$ (i.e., δ is unknown) $B = I$ is a satisfactory choice (which leads to the expressions in the previous section).
- The second condition $\mathrm{r}(XB) = \mathrm{r}(X',\Delta) - \mathrm{r}(\Delta)$ for determining B is equivalent to the condition $\mathrm{r}(X_dB) = \mathrm{r}(X_d',\Delta) - \mathrm{r}(\Delta)$.

Note that forming the $\mathbb{C}_d$ matrix as in (4.1) is consistent with forming the $\mathbb{C}$ matrix in the following two ways. First, the necessary and sufficient condition for H: $\Lambda'\delta = 0$ to be a linear hypothesis on δ is that $\Lambda'\delta$ is estimable and this condition is the same under the Y_d-model and the Y- model. Second, the choice of the matrix B was determined by the Y-model, but by the last observation above it could have equivalently been defined via the Y_d-model.

Lemma 4.1. Suppose that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ and that B is any matrix defined as in the definition of $\mathbb{C}$ or $\mathbb{C}_d$. Then

$$\mathrm{R}(\Lambda'B) = \mathrm{R}(\mathbb{C}) = \mathrm{R}(\mathbb{C}_d) = \{\Lambda'\delta : \delta \in \mathrm{N}(\Delta')\}.$$

Further, set $H = XB$ and $H_d = X_dB$. Then $\mathrm{E}(Y) = H\theta$, θ unknown, is a parametrization for $\mathrm{E}(Y)$ and $\mathrm{E}(Y_d) = H_d\theta$, θ unknown, is a parametrization for $\mathrm{E}(Y_d)$. Moreover, with respect to these parametrizations, the following statements can be made:

(a) If $\widehat{\theta}$ is *gm* for θ in the Y_d model (Y model), then $\widehat{\delta} = B\widehat{\theta}$ is *gm* for δ in the Y_d model (Y model).

(b) If $\pi'\delta$ is estimable, then $\pi'\mathrm{B}\theta$ is estimable and it's *blue* is the *blue* for $\pi'\delta$.

(c) H: $\Lambda' B\theta = 0$ is a linear hypothesis on θ that is equivalent to the linear hypothesis H: $\Lambda'\delta = 0$ in both the Y and Y_d models. That is,

$$\{X_d\delta\colon \Lambda'\delta = 0, \Delta'\delta = 0\} = \{H_d\theta\colon \Lambda' B\theta = 0\}$$

and the same equality is true with the sub d's removed.

Proof. Known linear model facts.

In a previous paragraph, the uniqueness of SSD_d was discussed. It seems appropriate to add a few additional comments to that discussion regarding the choice of Y_d. Note that SSD_d depends on $\Lambda'\widehat{\delta}$ and $\mathbb{C}_d$. Since $\Lambda'\widehat{\delta}$ is the *blue* for $\Lambda'\delta$, this part of SSD_d is unique and does not depend on the choice of Y_d. However, since $\mathbb{C}_d$ can be thought of as the covariance matrix of $\Lambda'\widehat{\delta}$ computed under the artificial assumption of a $\sigma^2 I$ covariance structure for Y_d, this part of SSD_d does depend on Y_d. Thus, SSD_d is uniquely defined for a given Y_d, but for a different choice of Y_d the associated SSD_d would be different. However, this difference is only through the matrix $\mathbb{C}_d$ and not through $\Lambda'\widehat{\delta}$.

4.3. Consistency of the devSS for the UM Procedure.

For testing a linear hypothesis in the usual linear model the devSS is invariant in the sense that no matter how the linear hypothesis is specified or computed one always arrives at the same devSS. That is, a linear hypothesis can be specified in different parametrizations and/or computed in seemingly different ways, but the end result is always the same. It is the purpose of this section to verify that the same thing is true for the UM devSS. Also, we establish that any g-inverse, as opposed to the Moore-Penrose inverse, can be used in computing $\mathbb{C}_d$ and SSD_d.

Consider the same setup as in the previous section. In particular, we have the linear model structure

$$\mathrm{E}(Y) = X\delta,\ \Delta'\delta = 0,\ \text{ and }\ \mathrm{Cov}(Y) = \sigma^2 I\ .$$

Further, X_d and T are matrices such that $X = TX_d$, T is $n \times d$ with rank d, $Y_d = (T'T)^{-1}T'Y$, $\mathrm{E}(Y_d) = X_d\delta$, $\Delta'\delta = 0$, and $\mathrm{Cov}(Y_d) = \sigma^2 V$ where $V = (T'T)^{-1}$. Also, assume that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ and that SSD_d is the UM devSS as described in (4.1). We first consider the question of computing SSD_d via an arbitrary g-inverse.

Lemma. The quantities $\mathbb{C}_d$ and SSD_d can be computed using arbitrary g-inverses instead of Moore-Penrose inverses.

Proof. Let G^- denote an arbitrary g-inverse of G. Let

$$\mathbb{A} = \Lambda' B(B'X_d'X_dB)^- B'\Lambda.$$

By estimability there exists A and L such that $X_d'A + \Delta L = \Lambda$. So,

$$\mathbb{A} = A'X_dB(B'X_d'X_dB)^- B'X_d'A = A'P_{X_dB}A.$$

Since the orthogonal projection operator P_{X_dB} on $\mathrm{R}(X_dB)$ is unique and since $\mathbb{C}_d$ can be expressed in exactly the same way, it follows that $\mathbb{A} = \mathbb{C}_d$. Now let $\mathrm{SS} = \widehat{\delta}\,'\Lambda\mathbb{C}_d^-\Lambda'\widehat{\delta}$. We want to show that $\mathrm{SS} = \mathrm{SSD}_d$ where SSD_d is defined the same as SS except with a Moore-Penrose inverse. Using Lemma 4.1(a) we can write $\Lambda'\widehat{\delta} = \Lambda' B\widehat{\theta}$ where $\widehat{\theta}$ is defined as in the lemma. Then using the above expression for $\mathbb{C}_d$ and $B'X_d'A = B'\Lambda$ also as above, we can write

$$\begin{aligned}\mathrm{SS} &= \widehat{\theta}B'X_d'A(A'P_{X_dB}A)^-A'X_dB\widehat{\theta}\\ &= \widehat{\theta}B'X_d'P_{X_dB}A(A'P_{X_dB}A)^-A'P_{X_dB}X_dB\widehat{\theta}\\ &= \widehat{\theta}B'X_d'FX_dB\widehat{\theta} \ = \widehat{\delta}\,'X_d'FX_d\widehat{\delta}\,,\end{aligned}$$

where F is the orthogonal projection operator on $\mathrm{R}(P_{X_dB}A)$. Since orthogonal projection operators are unique and since SSD_d can be expressed in exactly the same way as SS, it follows that $\mathrm{SS} = \mathrm{SSD}_d$.□

Now let us consider the question of computing SSD_d via different parametrizations. Suppose that another parametrization $\mathrm{E}(Y_d) = H_d\theta, \Gamma'\theta = 0$, is available and that H: $\pi'\theta = 0$ is a linear hypothesis on θ that is equivalent to H: $\Lambda'\delta = 0$. That is,

$$\{X_\mathrm{d}\delta\colon \Lambda'\delta = 0, \Delta'\delta = 0\} = \{H_\mathrm{d}\theta\colon \pi'\theta = 0, \Gamma'\theta = 0\}\ . \qquad (4.2)$$

Note that if $\mathrm{E}(Y) = H\theta$, $\Gamma'\theta = 0$, is a parametrization for $\mathrm{E}(Y)$ and if H: $\pi'\theta = 0$ is a linear hypothesis on θ that is equivalent to H: $\Lambda'\delta = 0$ (i.e., (4.2) is true with the sub d removed from X and H), then setting $H_d = (T'T)^{-1}T'H$ gives the setup described above. Now let SSD_h denote the UM devSS computed for the linear hypothesis H: $\pi'\theta = 0$ on θ. For future reference let us describe SSD_h. Following the description of expression (4.1) and using the comments in Section 4.2, let A be such that $\Gamma'A = 0$ and $\mathrm{r}(H_dA) = \mathrm{r}(H_d,\Gamma) - \mathrm{r}(\Gamma)$. Then

$$\mathrm{SSD}_h = (\pi'\widehat{\theta})'\mathbb{C}_h^+\pi'\widehat{\theta}\ ,$$

where $\mathbb{C}_h = \pi'A(A'H_d'H_dA)^+A'\pi$.

The remainder of this section is devoted to establishing that $\mathrm{SSD}_d = \mathrm{SSD}_h$. We do this in several parts, beginning with the most elementary situation of unconstrained models, i.e., Δ and Γ zero, and eventually establish the general result.

Lemma 4.2. Consider the case where $H_d = X_d$ and Δ, Γ are both zero. If (4.2) is true, then $\mathrm{R}(\Lambda) = \mathrm{R}(\pi)$ and $\mathrm{SSD}_d = \mathrm{SSD}_h$.

Proof. By estimability write $\pi = X_d'G$ and let $\overline{\delta} \in \mathrm{N}(\Lambda')$. Then (4.2) implies

$$X_\mathrm{d}\overline{\delta} = X_\mathrm{d}\overline{\overline{\delta}} \text{ for some } \pi'\overline{\overline{\delta}} = 0.$$

So, $X_{\mathrm{d}}(\bar{\delta} - \bar{\bar{\delta}}) = 0 \Rightarrow \pi'(\bar{\delta} - \bar{\bar{\delta}}) = A'X_{\mathrm{d}}(\bar{\delta} - \bar{\bar{\delta}}) = 0 \Rightarrow \bar{\delta} \in \mathrm{N}(\pi')$.

Thus, $\mathrm{N}(\Lambda') \subset \mathrm{N}(\pi')$. By a similar argument $\mathrm{N}(\pi') \subset \mathrm{N}(\Lambda')$. But this implies that $\mathrm{R}(\Lambda) = \mathrm{R}(\pi)$. Now let $D = (X_{\mathrm{d}}'X_{\mathrm{d}})^{+}$. Note that D is n.n.d. So,

$$\mathrm{r}(\Lambda' D\Lambda) = \mathrm{r}(D\Lambda) = \mathrm{r}(\Lambda) - \dim[\mathrm{R}(\Lambda) \cap \mathrm{N}(D)] = \mathrm{r}(\Lambda),$$

where the last equality follows because $\mathrm{R}(\Lambda) \subset \mathrm{R}(D) \Rightarrow \mathrm{R}(\Lambda) \cap \mathrm{N}(D) = \{0\}$. Now note that $\mathbb{C}_d = \Lambda' D\Lambda$ and $\mathbb{C}_h = \pi' D\pi$. Thus, $\mathrm{R}(\Lambda) = \mathrm{R}(\pi)$. So Theorem 4.11.8 in Rao and Mitra (1971) implies that $\Lambda\mathbb{C}_d^{+}\Lambda' = \pi\mathbb{C}_h^{+}\pi'$, and from this expression it is easy to see that $\mathrm{SSD}_d = \mathrm{SSD}_h$.□

Lemma 4.3. Consider the case where Δ and Γ are both zero. If (4.2) is true, then $\mathrm{SSD}_d = \mathrm{SSD}_h$.

Proof. Because of the parametrization we have $\mathrm{R}(X_d) = \mathrm{R}(H_d)$. Thus, let G be such that $H_d = X_d G$. Also, let $\widehat{\theta}$ be *gm* for θ. It is known from standard linear model facts that $\widehat{\delta} = G\widehat{\theta}$ is *gm* for δ and that H: $\Lambda'\delta = 0$ and H: $\Lambda' G\theta = 0$ are equivalent linear hypotheses. Also, by assumption H: $\Lambda'\delta = 0$ and H: $\pi'\theta = 0$ are equivalent linear hypotheses so that H: $\Lambda' G\theta = 0$ and H: $\pi'\theta = 0$ are equivalent linear hypotheses. Therefore, Lemma 4.2 implies that we can write

$$\mathrm{SSD}_h = \widehat{\theta}' G'\Lambda\mathbb{C}_h^{+}\Lambda' G\widehat{\theta} = \widehat{\delta}'\Lambda\mathbb{C}_h^{+}\Lambda'\widehat{\delta} \quad , \quad \mathbb{C}_h = \Lambda' G(H_{\mathrm{d}}' H_{\mathrm{d}})^{+} G'\Lambda$$

By estimability write $\Lambda = X_d' F$. Then this expression and $H_d = X_d G$ give

$$\begin{aligned}
\mathbb{C}_h &= F' X_d G(G' X_{\mathrm{d}}' X_{\mathrm{d}} G)^{+} G' X_d' F \\
&= F' P_{H_{\mathrm{d}}} F \\
&= F' P_{X_{\mathrm{d}}} F \;=\; F' X_{\mathrm{d}}(X_{\mathrm{d}}' X_{\mathrm{d}})^{+} X_{\mathrm{d}} F = \Lambda'(X_{\mathrm{d}}' X_{\mathrm{d}})^{+}\Lambda = \mathbb{C}_d.
\end{aligned}$$

This identity and the previous expression for SSD_h give the desired result.□

The previous two lemmas establish our contention of $\mathrm{SSD}_d = \mathrm{SSD}_h$ when both parametrizations are unconstrained. These lemmas are illustrated in the following example.

Example 4.1. Consider a 3×3 two-way model with interaction as described in model (2.1). Suppose that the incidence pattern and the cells means are

$$[n_{ij}] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 7 & 5 & 5 \end{bmatrix} \quad \text{and} \quad [Y_{ij.}] = \begin{bmatrix} 3 & 5 & 7 \\ 1 & 3 & 2 \\ 3 & 3 & 8 \end{bmatrix}.$$

Let us start by assuming the unconstrained parametrization $\mathrm{E}(Y) = X\delta$ described in model (2.1) with

$$\delta' = [\,\mu\ \alpha_1\ \alpha_2\ \alpha_3\ \beta_1\ \beta_2\ \beta_2\ \gamma_{11}\ \gamma_{12}\ \gamma_{13}\ \gamma_{21}\ \gamma_{22}\ \gamma_{23}\ \gamma_{31}\ \gamma_{32}\ \gamma_{33}].$$

Take X_d to be the matrix of distinct rows of X which gives

$$X_{\mathrm{d}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

The 27×9 matrix T is easily defined and the components of Y_d are the cell means given above. Also,

$$V = (T'T)^{-1} = \mathrm{diag}[\,1\,,\,1\,,\,1,\,1,\,\tfrac{1}{3}\,,\,\tfrac{1}{3}\,,\,\tfrac{1}{7}\,,\,\tfrac{1}{5}\,,\,\tfrac{1}{5}\,]\ .$$

Assume we want to test for no row effects as described in hypothesis HA3 of Table 2.5. In terms of the δ parameter this hypothesis can be written as

$$\mathrm{H}_1\colon \Lambda_1'\delta = 0$$

where

$$\Lambda_1' = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

Alternatively, it can also be written as

$$\mathrm{H}_2\!:\!\Lambda_2'\delta = 0$$

where $\Lambda_2' = \begin{bmatrix} 0 & 2 & -1 & -1 & 0 & 0 & 0 & \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}$

Note, as indicated in Lemma 4.3 that $R(\Lambda_1) = R(\Lambda_2)$. Now consider the cell mean parametrization with

$$X_{\mathrm{d}}^{\mu} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \delta_{\mu} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \end{bmatrix}$$

In terms of the cell mean parametrization, the hypothesis H_1 above can be expressed as

$$H_3\colon \Lambda_3'\delta_{\mu} = 0$$

where

$$\Lambda_3' = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}.$$

Also, consider another unconstrained full column parametrization $E(Y_d) = X_d^f\delta_f$ where

$$X_{\mathrm{d}}^{f} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \delta_f = \begin{bmatrix} \mu_{11}^f \\ \mu_{12}^f \\ \mu_{13}^f \\ \mu_{21}^f \\ \mu_{22}^f \\ \mu_{23}^f \\ \mu_{31}^f \\ \mu_{32}^f \\ \mu_{33}^f \end{bmatrix}.$$

In this parametrization we have that $\mu_{\mathrm{ij}}^{f} = \mu_{ij} - \mu_{3j}$ for $i = 1,2$ and $j = 1,2,3$ and $\mu_{3\mathrm{j}}^{f} = \mu_{3j}$ for $j = 1,2,3$. Using these equivalencies it is easy to write the H_3 linear hypothesis as

$$\mathrm{H}_4\text{: } \Lambda_4'\delta_f = 0$$

where

$$\Lambda_4' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

For the above four linear hypotheses, the $\mathbb{C}_d$ matrices are

$\mathbb{C}_{\mathrm{i}} = \Lambda_{\mathrm{i}}'(X_d'X_d)^{+}\Lambda_{\mathrm{i}}$, i = 1,2, $\mathbb{C}_3 = \Lambda_3'(X_{\mathrm{d}}^{\mu\prime}X_d^{\mu})^{-1}\Lambda_3$ and $\mathbb{C}_3 = \Lambda_3'(X_{\mathrm{d}}^{f\prime}X_d^{f})^{-1}\Lambda_3$

Thus, the UM devSS for these four hypotheses can be calculated as

$$\mathrm{SSD}_1 = \widehat{\delta}'\Lambda_1\mathbb{C}_1^{+}\Lambda_1'\widehat{\delta}, \text{ where } \widehat{\delta} = (X_{\mathrm{d}}'V^{-1}X_{\mathrm{d}})^{+}X_{\mathrm{d}}'V^{-1}Y_d,$$

$$\mathrm{SSD}_2 = \widehat{\delta}'\Lambda_2\mathbb{C}_2^{+}\Lambda_2'\widehat{\delta},$$

$$\mathrm{SSD}_3 = \widehat{\delta}_{\mu}'\Lambda_3\mathbb{C}_3^{+}\Lambda_3'\widehat{\delta}_{\mu}, \text{ where } \widehat{\delta}_{\mu} = (X_{\mathrm{d}}^{\mu\prime}V^{-1}X_{\mathrm{d}}^{f})^{-1}X_{\mathrm{d}}^{\mu\prime}V^{-1}Y_d,$$

$$\mathrm{SSD}_4 = \widehat{\delta}_{f}'\Lambda_4\mathbb{C}_4^{+}\Lambda_4'\widehat{\delta}_{f}, \text{ where } \widehat{\delta}_{f} = (X_{\mathrm{d}}^{f\prime}V^{-1}X_{\mathrm{d}}^{f})^{-1}X_{\mathrm{d}}^{f\prime}V^{-1}Y_d.$$

For the given data,

$\widehat{\delta}' = [$2.1875 1.5625 -0.6875 1.3125 -0.4375 0.5625 2.0625 -0.3125 0.6875
1.1875 -0.0625 0.9375 -1.5625 -0.0625 -1.0625 2.4375 $]$

$\widehat{\delta}_{\mu}' = [\,3\;\;5\;\;7\;\;1\;\;3\;\;2\;\;3\;\;3\;\;8\,]$

and $\widehat{\delta}_{f}' = [\,0\;\;2\;\;\text{-}1\;\;\text{-}2\;\;0\;\;\text{-}6\;\;3\;\;3\;\;8\,]$.

Using these calculations one can compute $\mathrm{SSD}_1 = \mathrm{SSD}_2 = \mathrm{SSD}\Lambda_3 = \mathrm{SS}_4 = 16.2222.\square$

The previous two lemmas and the above example are for unconstrained parametrizations. The next lemma extends Lemma 4.3 to constrained parametrizations.

Lemma 4.4. If (4.2) is true, then $\mathrm{SSD}_d = \mathrm{SSD}_h$.

Proof. Let $F = X_dB$ where B is the matrix in the definition of $\mathbb{C}_d$. Then Lemma 4.1 says that $\mathrm{E}(Y_d) = F\delta_1$, δ_1 unknown, is a parametrization for $\mathrm{E}(Y_d)$ and the linear hypothesis H_d: $\Lambda'\delta = 0$ on δ is equivalent to the linear hypothesis H_{d1}: $\Lambda' B\delta_1 = 0$ on δ_1. Let SSD_{d1} denote the UM devSS computed from the H_{d1} linear hypothesis. Using Lemma 4.1(b) is immediate to conclude that $\mathrm{SSD}_d = \mathrm{SSD}_{d1}$. Similarly form $U = H_d A$ where A is the matrix in the definition of SSD_h. In exactly the same manner as above conclude that $\mathrm{SSD}_h = \mathrm{SSD}_{h1}$ where SSD_{h1} is the UM devSS for H_{h1}: $\pi' A\theta_1 = 0$ relative to the parametrization $\mathrm{E}(Y_d) = U\theta_1$, θ_1 unknown. To conclude the proof note that (4.2) says H: $\Lambda'\delta = 0$ and H: $\pi'\theta = 0$ are equivalent linear hypotheses so that H_{d1} and H_{h1} are equivalent linear hypotheses. The conclusion now follows because Lemma 4.3 implies that $\mathrm{SSD}_{h1} = \mathrm{SSD}_{d1}$.□

Example 4.2. (Example 4.1 continued). Consider Example 4.1 with the constrained parametrization. That is,

$$\mathrm{E}(Y_{dij}) = \mu^c + \alpha_i^c + \beta_j^c + \gamma_{ij}^c\,,$$

where $i\,, j = 1,2,3$ and the parameters are constrained according to the following conditions

$$\Sigma_i\alpha_i^c = 0 = \Sigma_j\beta_j^c = 0 \quad \text{and} \quad \Sigma_i\gamma_{ij}^c = 0, \text{ all } i, \quad \text{and} \quad \Sigma_j\gamma_{ij}^c = 0\,, \text{ all } j\ .$$

Then this parametrization may be written as $\mathrm{E}(Y_d) = X_d\delta^c$, $\Delta'\delta^c = 0$, and $\mathrm{Cov}(Y_d) = \sigma^2 V$ where X_d and V are the same as in Example 4.1, δ^c has the same form as δ in Example 4.1, except that individual parameters having a superscript c, and Δ' is given by

$$\Delta' = \begin{bmatrix} 0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&1&1&1&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&1&1&1&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1\\ 0&0&0&0&0&0&0&1&0&0&1&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&1&0&0&1&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&1&0&0&1&0&0&1 \end{bmatrix}$$

Then from Section 2.3 a linear hypothesis equivalent to the linear hypotheses in Example 4.1 is given by H_5: $\alpha_1^c = \alpha_2^c = \alpha_3^c = 0$ or in matrix form as H_5: $\Lambda_5'\delta^c = 0$ where

$$\Lambda_5' = \begin{bmatrix} 0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0 \end{bmatrix}.$$

To get $\mathbb{C}_5$, the $\mathbb{C}_d$ matrix for this linear hypothesis, we can select any B satisfying $\mathrm{R}(B) = \mathrm{N}(\Delta_5')$, i.e., see the remarks in Section 4.3. One such B is

$$B = \begin{bmatrix} 1&0&0&0&0&0&0&0&0\\ 0&-1&-1&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0&0\\ 0&0&0&1&-1&0&0&0&0\\ 0&0&0&1&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&1&1&1\\ 0&0&0&0&0&-1&0&-1&0\\ 0&0&0&0&0&0&-1&0&-1\\ 0&0&0&0&0&-1&-1&0&0\\ 0&0&0&0&0&1&0&0&0\\ 0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&-1&-1\\ 0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&1 \end{bmatrix}.$$

Then $\mathbb{C}_5 = \Lambda_5' B(B'X_d'X_dB)^+B'\Lambda_5$. This B can also be used to get $\widehat{\delta}^c$ as follows: Set $Z_\mathrm{d} = X_\mathrm{d}B$. Then $\widehat{\delta}^c = B\widehat{\theta}$ where $\widehat{\theta} = (Z_\mathrm{d}'V^{-1}Z_\mathrm{d})^{-1}Z_\mathrm{d}'V^{-1}Y_d$.

Using the data in Example 4.1 we get

$$\widehat{\theta} = [3.8889\ \ -1.8889\ \ 0.7778\ \ -0.2222\ \ 1.7778\ \ 1.2222\ \ -1.7778\ -1.4444\ \ 1.5556']'$$

and $\widehat{\delta}^{c} = [3.8889\ \ 1.1111\ \ -1.8889\ \ 0.7778\ \ -1.5556\ \ -0.2222\ \ 1.7778\ \ -0.4444\ \ 0.2222$

$$0.2222\ \ 0.5556\ \ 1.2222\ \ -1.7778\ -0.1111\ \ -1.4444\ \ 1.5556]'\ .$$

And this leads to $\mathrm{SSD}_5 = \widehat{\delta}'\Lambda_5\mathbb{C}_5^{+}\Lambda_5'\widehat{\delta} = 16.2222$ as in Example 4.1.□

The previous two examples illustrate that SSD_d may be calculated in a number of different ways just as is well known for SSD.

4.4. Distribution of the devSS for the UM Procedure.

Suppose that Y is a random vector following the model in Section 4.2. Also, suppose that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ, that Y_d, $\mathbb{C}$, $\mathbb{C}_d$, SSD and $\mathrm{SSD_d}$ are defined as in Section 4.2 and that $\widehat{\delta}$ is gm for δ. It is the purpose of this section to explore some distributional facts about SSD_d. In this connection, we suppose throughout that Y has a MVN distribution.

First we need some notation. Suppose that $\mathbb{A}$ is a nonnegative definite matrix and that

$$\mathbb{A} = \textstyle\sum_i a_i A_i$$

is a spectral decomposition for $\mathbb{A}$ where the a_i are the positive eigenvalues of $\mathbb{A}$. Then $\mathbb{A}^{+} = \Sigma_i(1/a_i)A_i$, $\mathbb{A}^{\frac{1}{2}} = \Sigma_i\sqrt{a_i}A_i$, $\mathbb{A}^{-\frac{1}{2}} = \Sigma_i\sqrt{1/a_i}A_i$ and $\mathbb{A}^{-\frac{1}{2}}\mathbb{A}^{\frac{1}{2}} = \mathrm{P}_{\mathbb{A}}$. Also, let $\Omega_d = \{X_d\delta : \Delta'\delta = 0\}$ be the expectation space for the Y_d-model and let $\Omega_{\mathrm{H}d} = \{X_d\delta : \Delta'\delta = 0\ ,\ \Lambda'\delta = 0\}$ be the expectation space for the Y_d-model under the null hypothesis H: $\Lambda'\delta = 0$.

Lemma 4.5. $\mathrm{SSD_d}$ may be expressed in any of the following ways:

(a) $Z'\mathbb{C}_d^{+}\mathrm{Z}$ where $\mathrm{Z} = \Lambda'\widehat{\delta} \sim \mathrm{MVN}(\Lambda'\delta,\sigma^2\mathbb{C})$.

(b) $\mathrm{Z}'\mathbb{C}^{\frac{1}{2}}\mathbb{C}_d^{+}\mathbb{C}^{\frac{1}{2}}\mathrm{Z}$ where $\mathrm{Z} = \mathbb{C}^{-\frac{1}{2}}\Lambda'\widehat{\delta} \sim \mathrm{MVN}(\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta, \sigma^2\mathrm{P}_{\mathbb{C}})$

(c) $Y'\mathrm{P}_\Omega\mathrm{T}(\mathrm{T}'\mathrm{T})^{-1}F(\mathrm{T}'\mathrm{T})^{-1}\mathrm{T}'\mathrm{P}_\Omega\mathrm{Y}$ where F is the orthogonal projection operator on $\Omega_d \cap \Omega_{\mathrm{H}d}^{\perp}$ and P_Ω is the orthogonal projection operator on Ω.

Proof. Part (a) is immediate from the definition of SSD_d. Part (b) is easy using $\mathbb{C}^{\frac{1}{2}}\mathbb{C}^{-\frac{1}{2}} = \mathrm{P_C}$ and $\mathrm{P}_{\mathbb{C}}\Lambda'\widehat{\delta} = \Lambda'\widehat{\delta}$ because $\Lambda'\widehat{\delta}$ is proper. For part (c) we can use Lemmas 4.1 and 4.3 to suppose, without loss in generality, that $\Delta = 0$. By estimability, let A be such that $X_d'A = \Lambda$. Then,

$$\begin{aligned}
\mathrm{SSD}_d &= \widehat{\delta}X_d'A[A'X_d(X_d'X_d)^{+}X_d'A]^{+}A'X_d\widehat{\delta} \\
&= \widehat{\delta}X_d'P_{X_d}A[A'P_{X_d}A]^{+}A'P_{X_d}X_d\widehat{\delta} \\
&= \widehat{\delta}X_d'FX_d\widehat{\delta} \quad , \; F = (P_{X_d}A)[(P_{X_d}A)'(P_{X_d}A)]^{+}(P_{X_d}A)' \\
&= Y'X(X'X)^{+}X_d'FX_d(X'X)^{+}X'Y \\
&= Y'P_XT(T'T)^{-1}F(T'T)^{-1}T'P_XY \quad , X_d = (T'T)^{-1}T'X.
\end{aligned}$$

Note that $\Omega = \mathrm{R}(X)$ so that $\mathrm{P}_X = \mathrm{P}_\Omega$ and that F is the orthogonal projection operator on $S = \mathrm{R}(\mathrm{P}_{X_d}A)$. It remains to show that $\Omega_d \cap \Omega_{\mathrm{H}d}^{\perp} = S$. Clearly $S \subset \Omega_d$. Let $\mathrm{f} \in \Omega_{\mathrm{H}d}$. Write $\mathrm{f} = X_d\delta$, $\Lambda'\delta = 0$. Then

$$\mathrm{f}'\mathrm{P}_{X_d}A = \delta'X_d'\mathrm{P}_{X_d}A = \delta'X_d'A = \delta'\Lambda = 0 \;\Rightarrow\; S \subset \Omega_{\mathrm{H}d}^{\perp}.$$

So, $S \subset \Omega_d \cap \Omega_{\mathrm{H}d}^{\perp}$. To conclude the result we show that $\dim S$ is equal to

$$\dim[\Omega_d \cap \Omega_{\mathrm{H}d}^{\perp}] = \dim\Omega_d - \dim\Omega_{\mathrm{H}d} = \mathrm{r}(X_d) - [\mathrm{r}(X_d', \Lambda) - \mathrm{r}(\Lambda)] = \mathrm{r}(\Lambda).$$

Now note that $(X_d'X_d)^{+}$ is n.n.d. and its null space is $\mathrm{N}(X_d)$. So,

$$\begin{aligned}
\dim S &= \mathrm{r}[X_d(X_d'X_d)^{+}X_d'A] \\
&= \mathrm{r}[(X_d'X_d)^{+}\Lambda] = \mathrm{r}(\Lambda) - \dim[\mathrm{R}(\Lambda) \cap \mathrm{N}(X_d'X_d)^{+}] = \mathrm{r}(\Lambda)
\end{aligned}$$

because $\mathrm{R}(\Lambda) \subset \mathrm{R}(X_d')$.□

From (c) of the above lemma it is tempting to say that the previous section is unnecessary because for a given Y_d, SSD_d depends only on Ω_d and $\Omega_{\mathrm{H}d}$.

However, we used the results of the previous section to actually prove part (c). So, to omit the previous section the proof for part (c) would have to be changed.

In general SSD_d does not in general have an exact chi-squared distribution. However SSD_d can be expressed as a linear combination of chi-squared distributions.

Theorem 4.6. Set $\mathrm{Z} = \mathbb{C}^{-\frac{1}{2}}\Lambda'\widehat{\delta}$ and $H = \mathbb{C}^{\frac{1}{2}}\mathbb{C}_d^{+}\mathbb{C}^{\frac{1}{2}}$. Let $\pi_1,...,\pi_s$ be the distinct positive eigenvalues of H with multiplicities $\mathrm{r}_1,...,\mathrm{r}_s$ respectively and let $H = \Sigma_i\pi_i H_i$ be the spectral decomposition of H. Then

$$\mathrm{SSD}_d = \sum\nolimits_i \pi_i Q_i$$

where $Q_i = \mathrm{Z}'H_i\mathrm{Z} \sim \sigma^2\chi^2(\mathrm{r}_i, \lambda_i/\sigma^2)$ with $\lambda_i = \delta'\Lambda\mathbb{C}^{-\frac{1}{2}}H_i\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta$. Moreover, $Q_1, ..., Q_s$ are mutually independent.

Proof. Use Lemma 4.5(b) to write $\mathrm{SSD}_d = \Sigma_i\pi_i Q_i$. Recall that if A is symmetric, if $\mathrm{U} \sim \mathrm{MVN}(\theta,\Sigma)$ and if $\mathrm{A}\Sigma\mathrm{A} = \kappa\mathrm{A}$ where $\kappa > 0$, then $\mathrm{U}'\mathrm{AU}/\kappa \sim \chi^2(\underline{\mathrm{r}}(\mathrm{A}), \theta'\mathrm{A}\theta/\kappa)$. Because $\mathrm{P}_{\mathbb{C}} = H_1 + ... + H_s$, it is easy to check that $H_i\sigma^2 P_{\mathbb{C}} H_i = \sigma^2 H_i$ so that $Q_i \sim \sigma^2\chi^2(\mathrm{r}_i\, , \lambda_i/\sigma^2)$ with $\lambda_i = \delta'\Lambda\mathbb{C}^{-\frac{1}{2}}H_i\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta$. Moreover, $H_i P_{\mathbb{C}} H_j = 0$ for $i \neq j$ implies that $Q_1, ..., Q_s$ are mutually independent.□

Corollary 4.7. Consider the same notation and assumptions as in Theorem 4.6. Then the following statements can be made:

(a) $\mathrm{E}(\mathrm{SSD}_d) = \delta'\Lambda\mathbb{C}_d^{+}\Lambda'\delta + \sigma^2\mathrm{trace}(\mathbb{C}_d^{+}\mathbb{C})$.

(b) $\mathrm{E}(\mathrm{SSD}) = \delta'\Lambda\mathbb{C}^{+}\Lambda'\delta + \sigma^2\mathrm{r}(\mathbb{C})$.

(c) $\sum_i\lambda_i = \delta'\Lambda\mathbb{C}^{+}\Lambda'\delta$.

(d) $\delta'\Lambda\mathbb{C}_d^{+}\Lambda'\delta = \sum_i\pi_i\lambda_i$.

(e) $\mathrm{trace}(\mathbb{C}_d^{+}\mathbb{C}) = \sum_i \mathrm{r}_i\pi_i$.

(f) $\mathrm{SSD}_d/\sigma^2 \sim \sum_i \pi_i \chi^2(\mathrm{r}_i, \lambda_i/\sigma^2)$ where the chi-squareds are independent. Moreover, under the null hypothesis the λ_i are all zero and under the alternative hypothesis at least one λ_i is nonzero.

(g) $\mathrm{SSD}/\sigma^2 \sim \chi^2(\, \mathrm{r}(\mathbb{C})\, , \delta'\Lambda\mathbb{C}^+\Lambda'\delta/\sigma^2)$

Proof. Part (a) is easy using the expectation of a quadratic form and Lemma 4.5(a). Part (b) follows from (a) by substituting $\mathbb{C}^+$ for $\mathbb{C}_d^+$ and noting that trace$(\mathbb{C}^+\mathbb{C}) = \mathrm{r}(\mathbb{C})$. For (c) use Theorem 4.6 to write

$$\Sigma_i\lambda_i = \Sigma_i\delta'\Lambda\mathbb{C}^{-\frac{1}{2}}H_i\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta = \delta'\Lambda\mathbb{C}^{-\frac{1}{2}}\mathrm{P}_{\mathbb{C}}\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta$$

$$= \delta'\Lambda\mathbb{C}^{-\frac{1}{2}}\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta = \delta'\Lambda\mathbb{C}^+\Lambda'\delta \; .$$

For (d) note that

$$\begin{aligned}\Sigma_i\pi_i\lambda_i &= \Sigma_i\, \pi_i\delta'\Lambda\mathbb{C}^{-\frac{1}{2}}H_i\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta \\ &= \delta'\Lambda\mathbb{C}^{-\frac{1}{2}}H\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta \quad (\, H = \Sigma_i\pi_i H_i\,) \\ &= \delta'\Lambda\mathbb{C}^{-\frac{1}{2}}\mathbb{C}^{\frac{1}{2}}\mathbb{C}_d^+\mathbb{C}^{\frac{1}{2}}\mathbb{C}^{-\frac{1}{2}}\Lambda'\delta = \delta'\Lambda\mathbb{C}_d^+\Lambda'\delta \; .\end{aligned}$$

For (e) note that trace$(\mathbb{C}_d^+\mathbb{C}) =$ trace$(\mathbb{C}^{\frac{1}{2}}\mathbb{C}_d^+\mathbb{C}^{\frac{1}{2}}) =$ trace$(H) = \Sigma_i \mathrm{r}_i\pi_i$. The distribution part of (f) follows from Theorem 4.6 which also implies that under the null hypothesis $\lambda_i/\sigma^2 = 0$ because $\Lambda'\delta = 0$. Now suppose under the alternative hypothesis that $\lambda_i = 0$ for all i. Using the fact that $\mathbb{C}$ is n.n.d. and (c) we get

$$0 = \Sigma_i\lambda_i = \delta'\Lambda\mathbb{C}^+\Lambda'\delta \Rightarrow \Lambda'\delta \in \mathrm{N}(\mathbb{C}^+) = \mathrm{N}(\mathbb{C}).$$

Because $\Delta'\delta = 0$, Lemma 4.1 implies $\Lambda'\delta \in \mathrm{R}(\mathbb{C})$. So, $\Lambda'\delta = 0$. But this contradicts the alternative hypothesis assumption. Part (f) is well known or can be deduced from Theorem 4.6 with $H = \mathbb{C}^{\frac{1}{2}}\mathbb{C}^+\mathbb{C}^{\frac{1}{2}} = \mathrm{P}_{\mathbb{C}}$ which has a single eigenvalue of one with multiplicity $\mathrm{r}(\mathbb{C})$. □

Part (f) of this corollary essentially provides the relevant information to construct a test for a linear hypothesis based on the UM devSS. In particular, aside from dealing with σ^2, part (f) will imply that the test statistic has the same distribution for all distributions under the null hypothesis and will also imply that the statistic has a different distribution under the alternative hypothesis. These statements are more fully investigated in the next section.

4.5. The UM Test Statistic.

Let us consider the same model and notation as in Section 4.2. Also, suppose that H: $\Lambda'\delta = 0$ is a linear hypothesis on δ, that Y_d, $\mathbb{C}$, $\mathbb{C}_d$, SSD and $\mathrm{SSD_d}$ are defined as in Section 4.2 , that $\widehat{\delta}$ is gm for δ and that $V = (T'T)^{-1}$. It is the purpose of this section to introduce and explore the UM test statistic, i.e., the test based on SSD_d.

The test statistic using SSD is well known. For SSD_d we construct the test statistic in a similar fashion. Recall from Section 3.4 that the residual sum of squares SSE has $f = n - \dim\Omega$ degrees of freedom. Let $\mathrm{MSE} = \mathrm{SSE}/f$. Then the usual test statistic and the UM test statistic are:

$$\mathcal{F} = \mathrm{MSD}/\mathrm{MSE}\,, \qquad \mathrm{MSD} = \mathrm{SSD}/\mathrm{trace}(\mathbb{C}^{+}\mathbb{C})$$

$$\mathcal{F}_{\mathrm{d}} = \mathrm{MSD_d}/\mathrm{MSE}\,, \quad \mathrm{MSD_d} = \mathrm{SSD_d}/\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C})$$

The divisors for the mean squares are determined so that their expectations are of the form σ^2 plus a constant. The expected mean squares are

$$\mathrm{E(MSD)} = \sigma^2 + \varphi\,, \qquad \varphi = \frac{\delta'\Lambda\mathbb{C}^{+}\Lambda'\delta}{\mathrm{trace}(\mathbb{C}^{+}\mathbb{C})}$$

$$\mathrm{E(MSD}_d) = \sigma^2 + \varphi_d\,, \qquad \varphi_d = \frac{\delta'\Lambda\mathbb{C}_{\mathrm{d}}^{+}\Lambda'\delta}{\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C})}$$

The quantity φ is frequently referred to as the noncentrality parameter. We shall also use this term for φ_d even though this is somewhat misleading. Some basic information about the mean squares and their expectation is given in the next lemma.

Lemma 4.8. The following three statements are equivalent:

$$\varphi = 0\ , \quad \varphi_d = 0 \ \text{ and } \ \Lambda'\delta = 0\ .$$

Furthermore, MSD_d and MSE are independent.

Proof. If $\Lambda'\delta = 0$, then clearly $\varphi = 0$. Conversely, $\varphi = 0$ implies that $\Lambda'\delta \in \mathrm{N}(\mathbb{C}^+) = \mathrm{N}(\mathbb{C})$. But $\Delta'\delta = 0$ so that Lemma 4.1 implies that $\Lambda'\delta$ is also in $\mathrm{R}(\mathbb{C})$. Thus, $\Lambda'\delta = 0$. The proof for φ_d is essentially identical. The independence follows because $\mathrm{MSD_d}$ is a function of $\Lambda'\widehat{\delta}$ which is independent of MSE.□

Now let us examine the test statistic $\mathcal{F}_d$. By using Proposition 3.5(a) and Corollary 4.7(c, f) it is clear that the distribution of $\mathcal{F}_d$ can be expressed as

$$\mathcal{F}_d \sim \frac{\Sigma_i c_i \chi^2(\mathrm{r}_i, \lambda_i/\sigma^2)}{\chi^2(f)/f}\ , \quad c_i = \pi_i/\Sigma_i \mathrm{r}_i \pi_i\ , \tag{4.4}$$

where all of the chi-squareds are independent and the λ_i, π_i are defined as in Theorem 4.6. Since the λ_i are all zero when $\Lambda'\delta = 0$, it is clear from this expression that $\mathcal{F}_d$ has a single null distribution. That is, $\mathcal{F}_d$ has the same distribution for all distributions under the null hypothesis and its distribution does not depend on any unknown parameters. Thus, a critical value can be computed for $\mathcal{F}_d$ without confusion. Let K denote the critical value for the usual test and let K_d denote the critical value for $\mathcal{F}_d$. That is,

$$\Pr\{\mathcal{F} \geq \mathrm{K} | \mathrm{H} \text{ is true}\} = \Pr\{\mathcal{F}_d \geq \mathrm{K}_d | \mathrm{H} \text{ is true}\} = \alpha$$

where α is some fixed level of significance. Next we wish to examine the test under the alternative. For the distribution under the alternative, Corollary 4.7(f) implies at least one of the λ_i is nonzero so that the distribution of $\mathcal{F}_d$ under the alternative is not the same as the null distribution. This is sufficient to conclude that the test using $\mathcal{F}_d$ is unbiased. To actually prove this we need the following lemma which a generalization of known results for the usual test statistic $\mathcal{F}$, i.e., the special case $Z = 0$ and $Q = 1$ reduces to the usual result for non-central chi-squared random variables.

Lemma 4.9. Suppose that Z and $Q > 0$ are random variables such that $\Pr\{Z/Q < x\} > 0$ for all positive x. For each positive integer k and each $\lambda \geq 0$, let

$$h(k,\lambda) = \Pr\{\ (c\chi^2(k,\lambda) + Z)/Q \geq x\ \}\ ,$$

where c and x are positive real numbers. Assume that Z and Q are jointly independent of $\chi^2(k,\lambda)$. Then $h(k,\lambda)$ is a strictly increasing function of k and a strictly increasing function of λ.

Proof. Let $H = h(k,\lambda^*) - h(k,\lambda)$ where $\lambda^* > \lambda \geq 0$. By using conditional expectation we can write

$$h(k,\lambda) = E_{Z,Q}\Big[\Pr\{(c\chi^2(k,\lambda) + Z)/Q \geq x\ |Z,Q\}\Big],$$

and because of the independence we can write

$$\Pr\{(c\chi^2(k,\lambda) + Z)/Q \geq x\ |Z = z, Q = q\} = \Pr\{\chi^2(k,\lambda) \geq \tfrac{qx-z}{c}\ \}.$$

Since these expressions can also be written with λ^* in place of λ we can write $H = E[G(Z,Q)]$ where

$$g(z,q) = \Pr\{\chi^2(k,\lambda^*) \geq \tfrac{qx-z}{c}\ \} - \Pr\{\chi^2(k,\lambda) \geq \tfrac{qx-z}{c}\ \}.$$

From standard linear model theory we know that $g(z,q) \geq 0$ and is strictly positive if $qx - z > 0$. Thus, $H \geq 0$ and if $H = 0$, then (24) in Section 1.2 of Lehmann (1983) implies that $\Pr\{Z/Q \leq x\} = 0$ which is a contradiction. A similar proof works for $H = h(k^*,\lambda) - h(k,\lambda)$ where $k^* > k$. □

Proposition 4.10. The power function for the test statistic $\mathcal{F}_d$ is strictly increasing in each of the noncentrality parameters $\lambda_1,\ldots,\lambda_s$.

Proof. For $\mathcal{F}_d$ use the expression in (4.4). Then use Lemma 4.9 with $k = r_1$, $\lambda = \lambda_1/\sigma^2$, $c = c_1$, $Q = MSE$, and $Z = c_2Q_1 + \ldots + c\pi_sQ_s$. To apply the lemma simply note that the condition $\Pr\{Z/Q < x\} > 0$ is true for all positive x and that the independence condition is also satisfied. The remaining λ_i and r_i can be treated in a similar manner.□

Notice that Proposition 4.10 and Corollary 4.7(f) imply that the power of the UM test under any alternative is greater than the significance level so that the UM test is unbiased. In fact, it is stronger than unbiased since the power function is strictly increasing in each noncentrality parameter λ_i.

In general we would like to compare the tests based on $\mathcal{F}$ and $\mathcal{F}_d$. The usual test is based on an F-distribution with $r = \text{rank}(\mathbb{C})$ degrees of freedom. By using $\Sigma_i r_i = r$, Corollary 4.7(c, g) and the additive property of the chi-squared distribution we can write

$$\mathcal{F} \sim \frac{a\chi^2(r, r\varphi/\sigma^2)}{\chi^2(f/f} \sim \frac{\Sigma_i a_i \chi^2(r_i, \lambda_i/\sigma^2)}{\chi^2(f)/f} \quad , \quad a = a_i = 1/r\,, \tag{4.5}$$

where all of the chi-squared variables are independent. In this statement the first expression is the usual expression for $\mathcal{F}$ and the second expression is written so that a more direct comparison with $\mathcal{F}_d$ is available. Notice in both expressions (4.4) and (4.5) that the constants c_i and the constants a_i add to one if the multiplicities are all equal to one. This summation of the constants to one can also be achieved in two other ways. First, we could suppose that r_i are all equal to one, but then in Theorem 4.6 there would be r not necessarily distinct eigenvalues. Secondly, one could use the constants $r_i c_i$ and $r_i a_i$ with mean squares instead of chi-squareds in the numerators, i.e., divide the ith chi-squared by r_i.

The next example illustrated the ideas presented above. Before giving the example we state, without proof, some linear algebra facts in the following lemmas.

Lemma 4.11. Suppose that $H = \mathbb{C}^{\frac{1}{2}}\mathbb{C}_d^{+}\mathbb{C}^{\frac{1}{2}}$ as well as the π_i, r_i, H_i and λ_i are defined as in Theorem 4.6. Let $A = \mathbb{C}_d^{+}\mathbb{C}$. Then the following statements may be made:

(a) If $HG = \pi G$, then $AF = \pi F$ where $F = \mathbb{C}^{-\frac{1}{2}}G$.

(b) If $AF = \pi F$, then $HG = \pi G$ where $G = \mathbb{C}^{\frac{1}{2}}F$.

(c) H and A share the same eigenvalues and multiplicities.

(d) If $AF = \pi_i F$ and $\mathrm{r}(F) = \mathrm{r}_i$, then

$H_i = \mathbb{C}^{\frac{1}{2}}F(F'\mathbb{C}F)^{+}F'\mathbb{C}^{\frac{1}{2}}$ and $\lambda_i = \delta'\Lambda F(F'\mathbb{C}F)^{+}F'\Lambda'\delta$.

Notice that this lemma allows us to calculate the quantities needed to describe the distribution characteristics in Theorem 4.6 without computing any quantities with the $\frac{1}{2}$ or $-\frac{1}{2}$ exponents.

Example 4.4. (Two-way model with interaction, continued). Consider Model (2.1) with $a = 3$ rows and b columns and with the cell mean parametrization. Assume interest is in testing for HA3 (or equivalently HAU) in Table 2.5. That is, for "no row effects" via SAS type III. Then our interest is in testing H: $\theta_i = \theta_j$ for all i, j where θ_i is the average of the cell means in row i, i.e., $\theta_i = (\Sigma_j \mu_{ij})/b$. Let us begin by finding $\mathbb{C}$ and $\mathbb{C}_d$. Instead of using the expressions in Section 4.2 it is easier to compute the covariance matrices directly. Let $\theta = (\theta_1,...,\theta_3)'$, let $\widehat{\theta}$ be the *blue* for θ and let D be such that $\mathrm{Cov}(\widehat{\theta}) = \sigma^2\mathrm{D}$. Since $\widehat{\mu}_{ij} = \overline{Y}_{ij.}$ note that

$$\mathrm{D} = b^{-1}\mathrm{diag}(d_1,...,d_3) \quad , \quad d_i = (\Sigma_j 1/n_{ij})/b \ .$$

Let us write the null hypothesis as H: $\Lambda'\delta = 0$ where

$$\Lambda' = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & -2 \end{bmatrix}.$$

Then the null hypothesis of interest can be expressed as H: $\Lambda'\theta = 0$. Using this expression and the previous information we can write

$$\mathbb{C} = \Lambda'\mathrm{D}\Lambda = b^{-1}\begin{bmatrix} d_1 + d_2 & d_1 - d_2 \\ d_1 - d_2 & d_1 + d_2 + 4d_3 \end{bmatrix} \text{ and } \mathbb{C}_d = \Lambda'\Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}.$$

Let us compute φ and φ_d. First note that $\Lambda\mathbb{C}_d^+\Lambda' = \Lambda(\Lambda'\Lambda^+)\Lambda' = P_\Lambda$. since $\Lambda'\mathbf{1} = 0$ it follows that $P_\Lambda = (I - P_1)$. Thus, φ_d can be computed via the ratio of

$$\theta'\Lambda\mathbb{C}_d^+\Lambda'\theta = \theta'(I - P_1)\theta \quad \text{and} \quad \mathrm{trace}(\mathbb{C}_d^{-1}\mathbb{C}) = 2(d_1 + d_2 + d_3)/3b$$

Next, note that

$$\Lambda\mathbb{C}^+\Lambda' = \kappa\begin{bmatrix} d_2 + d_3 & -d_3 & -d_2 \\ -d_3 & d_1 + d_3 & -d_1 \\ -d_2 & -d_1 & d_1 + d_2 \end{bmatrix}, \ \kappa = \frac{b}{d_1d_2 + d_1d_3 + d_2d_3}.$$

Since $\mathrm{trace}(\mathbb{C}^{-1}\mathbb{C}) = 2$, note that φ can easily be computed. Now set

$$A = 6b\mathbb{C}_d^{-1}\mathbb{C} = \begin{bmatrix} 3(d_1 + d_2) & 3(d_1 - d_2) \\ (d_1 - d_2) & d_1 + d_2 + 4d_3 \end{bmatrix} = \begin{bmatrix} a & 3m \\ m & g \end{bmatrix}.$$

Note that the normalized eigenvalues of A (that is, the eigenvalues of A divided by the trace of A) are the same as the c_i in (4.4). Let x_i be the associated eigenvectors. If $m = 0$, then

$$c_1 = a/(a+g) \text{ with } x_1 = [1, 0]'$$

and
$$c_2 = g/(a+g) \text{ with } x_2 = [0, 1]'.$$

If $m \neq 0$, then the quadratic equation and the fact that $A - \pi I$ has rank one gives

$$c_i = \tfrac{1}{2}\left((a+g) \pm \sqrt{(a-g)^2 + 12m^2}\right)/(a+g)$$

with $x_i = (g - c_i(a+g), -m)$. Using Lemma 4.11, we can easily calculate noncentrality parameters, especially if $d_1 = d_2$ so that $m = 0$. In this case

$$\lambda_1 = \theta'\Lambda x_1(x_1'\mathbb{C}x_1)^+ x_1'\Lambda'\theta$$

$$= [\theta_1\,\theta_2\,\theta_3]\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\left(b^{-1}[1\ \ 0]\begin{bmatrix} d_1 + d_2 & d_1 - d_2 \\ d_1 - d_2 & d_1 + d_2 + 4d_3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)^+[1\ \ 0]\begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & -2 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

$$= \frac{b(\theta_1 - \theta_2)^2}{d_1 + d_2} = \frac{b(\theta_1 - \theta_2)^2}{2d_1}$$

$\lambda_2 = \theta'\Lambda\, x_2(\, x_2'\mathbb{C}\, x_2)^+\, x_2'\Lambda'\theta = \frac{b(\theta_1+\theta_2-2\theta_3)^2}{d_1+d_2+4d_3} = \frac{b(\theta_1+\theta_2-2\theta_3)^2}{2(d_1+2d_3)}.$

Table (4.1) below illustrates a some specific computations for the quantities discussed in this example for a number of different n_{ij} patterns. In all of the examples we have $a = b = 3$. Also, the designs are identified by the d_i instead of the incidence pattern. For example, if $d_i = 1$, then we must have $n_{ij} = 1$ for $j = 1,2,3$. However, other d_i might be come from different n_{ij} patterns. For example, consider the first line where $d_1 = d_2 = 1$ and $d_3 = .1$. This d_i pattern could be generated by any of the following n_{ij} patterns:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 10 & 10 & 10 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 5 & 20 & 20 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 20 & 5 & 20 \end{bmatrix}$$

It is also interesting to note that if a specific n_{ij} pattern leads to the d_i in the table, then one can also compute the table n_{ij} pattern given by $\mathrm{k}n_{ij}$ for all i, j. In particular $\mathrm{k}n_{ij}$ leads to the same c_i with the remaining quantities λ_1, λ_2, φ, and φ_d all being multiplied by k and the d_i being divided by k. For example, suppose one has the n_{ij} pattern

$$\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 20 & 20 & 20 \end{bmatrix}.$$

Since this is twice the incidence pattern of the first line in the table one could compute the corresponding table values for this incidence pattern as

$d_1 = d_3 = .5$, $d_3 = .05$, $c_1 = 0.7143$, $c_2 = 0.2857$

$\lambda_1 = 3.0$, $\lambda_2 = 2.5000$, $\varphi_d = 2.8572$ and $\varphi = 2.75$.

Above we discussed how Table 4.1 can give additional information for certain other n_{ij} patterns. If is also possible to use the table to get other n_{ij} patterns for

Table 4.1 Noncentral parameters and expected mean squares

θ'	d_1	d_2	d_3	c_1	c_2	λ_1	λ_2	φ_d	φ
[1 0 0]	1.0000	1.0000	0.1000	0.7143	0.2857	1.5000	1.2500	1.4286	1.3750
	1.0000	1.0000	0.0500	0.7317	0.2683	1.5000	1.3636	1.4634	1.4318
	1.0000	1.0000	0.0200	0.7426	0.2574	1.5000	1.4423	1.4851	1.4712
	1.0000	1.0000	0.0100	0.7463	0.2537	1.5000	1.4706	1.4925	1.4853
	1.0000	1.0000	0.0010	0.7496	0.2504	1.5000	1.4970	1.4993	1.4985
[0 0 1]	1.0000	1.0000	0.1000	0.7143	0.2857	0.0000	5.0000	2.5000	2.5000
	1.0000	1.0000	0.0500	0.7317	0.2683	0.0000	5.4545	2.7273	2.7273
	1.0000	1.0000	0.0200	0.7426	0.2574	0.0000	5.7692	2.8846	2.8846
	1.0000	1.0000	0.0100	0.7463	0.2537	0.0000	5.8824	2.9412	2.9412
	1.0000	1.0000	0.0010	0.7496	0.2504	0.0000	5.9880	2.9940	2.9940
[1 0 0]	1.0000	0.5000	0.1667	0.7179	0.2821	2.4039	0.2627	1.8000	1.3333
	1.0000	0.5556	0.1810	0.7045	0.2955	2.3169	0.3228	1.7276	1.3198
[0 0 1]	1.0000	0.5000	0.1667	0.7179	0.2821	0.2470	5.7530	1.8000	3.0000
	1.0000	0.5556	0.1810	0.7045	0.2955	0.1957	5.3795	1.7279	2.7876

different values of b. In particular, suppose you have an n_{ij} pattern for a $3 \times b$ design that gives the same d_i values as in the above table. Then the other quantities for the $3 \times b$ pattern can be computed from the above table as follows. The c_i are the same and $\lambda_1, \lambda_2, \varphi$ and φ_d are obtained from the above table by multiplying the above table quantities by $b/3$. For example, consider the following n_{ij} pattern

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 10 & 10 & 10 & 10 \end{bmatrix}.$$

For this pattern we have $b = 4$, $d_1 = d_2 = 1$ and $d_3 = 0.10$. Since these d_i values are the same as in the first line of the above table we can compute the other six quantities for this 3×4 n_{ij} pattern as follows: $c_1 = .7143$, $c_2 = .2857$, $\lambda_1 = (4/3)1.5 = 2$, $\lambda_2 = (4/3)1.25 = 1.66667$, $\varphi_d = (4/3)1.4286 = 1.9048$ and $\varphi = (4/3)1.375 = 1.8333.\Box$

Note that in the above example if the d_i are all equal to κ, then $\mathbb{C}_d = (b/\kappa)\mathbb{C}$ from which it is almost immediate to see that $\mathrm{MSD} = \mathrm{MSD}_d$. This condition is generalized in the following proposition.

Proposition 4.12. The following statements are quivalent:

(a) $\mathrm{MSD} = \mathrm{MSD}_d$.

(b) $\mathbb{C}_d$ is a constant times $\mathbb{C}$

(c) $\mathbb{C}_d^+\mathbb{C}$ has only one positive eigenvalue.

Proof. If (a) is true,then

$$\mathbb{C}^+/\mathrm{tr}(\mathbb{C}^+\mathbb{C}) = \mathbb{C}_d^+/\mathrm{tr}(\mathbb{C}_d^+\mathbb{C}).$$

So, $\mathbb{C}_d^+ = \kappa\mathbb{C}^+$ where $\kappa = \mathrm{tr}(\mathbb{C}_d^+\mathbb{C})/\mathrm{tr}(\mathbb{C}^+\mathbb{C})$. Thus, (a) $\Rightarrow$ (b). Now suppose that (b) is true. Then $\mathbb{C}_d = \kappa\mathbb{C}$. So, $\mathbb{C}_d^+\mathbb{C} = \kappa^{-1}\mathbb{C}^+\mathbb{C} = \kappa^{-1}\mathrm{P}_{\mathbb{C}}$ which has κ^{-1} as its only positive eigenvalue . Thus, (b) $\Rightarrow$ (c). Now suppose that (c) is true. Then $\mathbb{C}_d^+\mathbb{C} = \kappa\mathrm{P}_{\mathbb{C}}$ for some κ. So,

$$\mathbb{C} = \mathrm{P}_{\mathbb{C}}\mathbb{C} = \mathbb{C}_d\mathbb{C}_d^+\mathbb{C} = \kappa\mathbb{C}_d\mathrm{P}_{\mathbb{C}} = \kappa\mathbb{C}_d.$$

Thus,

$$\mathrm{MSD}_d = (\Lambda'\widehat{\delta})'\mathbb{C}_d^+(\Lambda'\widehat{\delta})/\mathrm{tr}(\mathbb{C}_d^+\mathbb{C}) = (\Lambda'\widehat{\delta})'[\kappa\mathbb{C}^+](\Lambda'\widehat{\delta})/[\kappa\mathrm{tr}(\mathbb{C}^+\mathbb{C})] = \mathrm{MSD}.$$

Therefore, (c) $\Rightarrow$ (a) which concludes the proof.

Corollary 4.13. In any of the following situations, $\mathrm{MSD} = \mathrm{MSD}_d$.

(a) Testing for no row effects in the two-way model with interaction when the d_i are all equal.

(b) $\mathrm{r}(\mathbb{C}) = 1$.

(c) Testing for any effect in a 2^k design.

Proof. Easy.□

The previous examples have illustrated some of the various calculations for determining the distribution of $\mathcal{F}_d$. Some authors like El-Bassiouni and El-Shahat 1986) and some of our calculations in the next chapter suggest that examining φ and φ_d can provide insight into the power of $\mathcal{F}$ and $\mathcal{F}_d$. Because of this we more fully explores φ and φ_d in the following example.

Example 4.5. Again consider the previous $3 \times b$ two-way model with interaction. Let us adopt the same notation as in Example 4.4 and again suppose we are interested in testing H: $\theta_i = \theta_j$ for all i, j. Then it can be established that

$$\varphi = a_1(\theta_1 - \theta_2)^2 + a_2(\theta_1 - \theta_3)^2 + a_3(\theta_2 - \theta_3)^2$$

where $a_1 = \frac{b}{(a-1)} \frac{d_3}{d_1 d_2 + d_1 d_3 + d_2 d_3}$,

$$a_2 = \frac{b}{(a-1)} \frac{d_2}{d_1 d_2 + d_1 d_3 + d_2 d_3} ,$$

$$a_3 = \frac{b}{(a-1)} \frac{d_1}{d_1 d_2 + d_1 d_3 + d_2 d_3} .$$

Also,

$$\varphi_d = \frac{bn_h}{a(a-1)}[(\theta_1 - \theta_2)^2 + (\theta_1 - \theta_3)^2 + (\theta_2 - \theta_3)^2]$$

where $n_h = 3/(d_1 + d_2 + d_2)$ is the harmonic mean of all the cell sizes. Note that in disproportional designs the a_i are in general not the same so that the weighting of the mean differences making up φ depend on the cell sizes. However, this is not the case in φ_d where each mean difference is given the same weight $bn_h/6$. To get an idea about the weights several different n_{ij} patterns are given in Table 4.2 along with the a_i and $bn_h/6$. Note in the first three n-patterns that all of the d_i are equal so that Corollary 4.13(a) implies that $\text{MSD} = \text{MSD}_d$.□

Table 4.2 Coefficients on expected mean squares

n-pattern	Methods	d_1	d_2	d_3	a_1	a_2	a_3
3 3 3	MSD(Type III)	0.3333	0.3333	0.3333	1.5000	1.5000	1.5000
3 3 3	MSD$_d$				1.5000	1.5000	1.5000
3 3 3							
1 2 6	MSD(Type III)	0.5556	0.5556	0.5556	0.9000	0.9000	0.9000
1 2 6	MSD$_d$				0.9000	0.9000	0.9000
1 2 6							
1 2 6	MSD(Type III)	0.5556	0.5556	0.5556	0.9000	0.9000	0.9000
6 1 2	MSD$_d$				0.9000	0.9000	0.9000
2 6 1							
1 1 1	MSD(Type III)	1.0000	0.5000	0.1667	0.3333	1.0000	2.0000
2 2 2	MSD$_d$				0.9000	0.9000	0.9000
6 6 6							
1 1 1	MSD(Type III)	1.0000	0.5556	0.1810	0.3243	0.9956	1.7920
1 3 3	MSD$_d$				0.8638	0.8638	0.8638
7 5 5							

4.6. An Alternative for SSD$_d$.

Consider the same model and notation as in previous section. Recall that SSD$_d$ has the form $\widehat{\delta}'\Lambda\mathbb{C}_d^+\Lambda'\widehat{\delta}$ where $\widehat{\delta}$ is *gm* for δ. An alternative is to use $\tilde{\delta}$ in place of $\widehat{\delta}$ where $\tilde{\delta}$ is least squares for δ based on the Y_d model. That is, $\tilde{\delta}$ is *gm* for δ based on the Y_d model computed as though "$\mathrm{Cov}(Y_d) = \sigma^2 I$" instead of the actual $\mathrm{Cov}(Y_d) = \sigma^2 V$. Note that the computation of $\tilde{\delta}$ is consistent with the computation of $\mathbb{C}_d$ and it is also consistent with the way the UM sums of squares have been interpreted in the UM literature on variance component confidence intervals, e.g., see Burdick and Graybill (1992, section 6.6). For purposes of comparison let SSD$_\mathrm{R}$ denote this devSS. That is,

$$\mathrm{SSD_R} = \tilde{\delta}'\Lambda'\mathbb{C}_d^+\Lambda\tilde{\delta}\,, \tag{4.6}$$

where $\mathbb{C}_d$ has its usual definition and $\tilde{\delta}$ is least squares for δ based on the Y_d model. Following as in the previous section the test statistic is

$$\mathcal{F}_{\mathrm{R}} = \mathrm{MSD}_{\mathrm{R}}/\mathrm{MSE}\,, \quad \mathrm{MSD}_{\mathrm{R}} = \mathrm{SSD}_{\mathrm{d}}/\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C}_{\mathrm{R}}),$$

and the expected mean square is

$$\mathrm{E}(\mathrm{MSD}_{\mathrm{R}}) = \sigma^2 + \varphi_{\mathrm{R}} \quad , \qquad \varphi_{\mathrm{R}} = \frac{\delta'\Lambda\mathbb{C}_{\mathrm{d}}^{+}\Lambda'\delta}{\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C}_{\mathrm{R}})}$$

Some basic information about $\Lambda'\tilde{\delta}$ is given in the following lemma.

Lemma 4.14. Let B be any matrix satisfying $\mathrm{r}(X_dB) = \mathrm{r}(X_d', \Delta) - \mathrm{r}(\Delta)$ and $\Delta'B = 0$. Then $\Lambda'\tilde{\delta} \sim \mathrm{MVN}(\Lambda'\delta, \sigma^2\mathbb{C}_{\mathrm{R}})$ where

$$\mathbb{C}_{\mathrm{R}} = L_{\mathrm{R}}'VL_{\mathrm{R}} \quad , \quad L_{\mathrm{R}} = X_{\mathrm{d}}B(B'X_{\mathrm{d}}'X_{\mathrm{d}}B)^{+}B'\Lambda\,,$$

Moreover,

(a) $\mathrm{E}(\Lambda'\hat{\delta}) = \mathrm{E}(\Lambda'\tilde{\delta})$.

(b) $\mathrm{Cov}(\Lambda'\hat{\delta}) \le \mathrm{Cov}(\Lambda'\tilde{\delta})$.

(c) $\mathrm{R}(VL_{\mathrm{R}}) \subset \mathrm{R}(X_dB) \Leftrightarrow \Lambda'\hat{\delta} = \Lambda'\tilde{\delta}$.

(d) $\varphi = 0 \Leftrightarrow \Lambda'\delta = 0$.

(e) $\varphi_{\mathrm{R}} \le \varphi_d$.

(f) $\mathrm{SSD}_{\mathrm{R}}$ and MSE are independent if and only if $\Lambda'\hat{\delta} = \Lambda'\tilde{\delta}$.

Proof. Part (a) is a known fact. Part (b) follows from (a) and the fact that $\Lambda'\hat{\delta}$ is the *blue* for $\Lambda'\delta$. Part (c) follows from Zyskind's Theorem, the uniqueness of *blues* and the observation that $\Lambda'\tilde{\delta} = L_{\mathrm{R}}'Y_d$. Part (d) follows from a proof almost identical to the proof for Lemma 4.8. For part (e) note that $\mathbb{C}_d^{+}$ is nonnegative definite. Thus, there exists A such that $\mathbb{C}_d^{+} = AA'$. From Lemma 4.5(a) we have $\mathrm{Cov}(\Lambda'\hat{\delta}) = \sigma^2\mathbb{C}$. Also, $\mathrm{Cov}(\Lambda'\tilde{\delta}) = \sigma^2\mathbb{C}_{\mathrm{R}}$. Thus,

$$(b) \Rightarrow \quad \mathbb{C} \le \mathbb{C}_{\mathrm{R}} \;\Rightarrow\; A\mathbb{C}A' \le A\mathbb{C}_{\mathrm{R}}A'$$

$$\Rightarrow \mathrm{trace}(A\mathbb{C}A') \le \mathrm{trace}(A\mathbb{C}_{\mathrm{R}}A')$$

$$\Rightarrow \mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C}) \le \mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C}_{\mathrm{R}})$$

$$\Rightarrow \frac{\delta'\Lambda\mathbb{C}_{\mathrm{d}}^{+}\Lambda'\delta}{\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C}_{\mathrm{R}})} \le \frac{\delta'\Lambda\mathbb{C}_{\mathrm{d}}^{+}\Lambda'\delta}{\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C})} \Rightarrow \varphi_{\mathrm{R}} \le \varphi_{d}.$$

For part (f), note that if $\Lambda'\widehat{\delta} = \Lambda'\tilde{\delta}$, then $\mathrm{SSD_R} = \mathrm{SSD}_d$ which is independent of MSE. Conversely, if $\mathrm{SSD_R}$ and MSE are independent, then it can be establsihed that the covariance between $\Lambda'\tilde{\delta}$ and the least squares residuals $(I - P_\Omega)Y$ is zero. But this means $\Lambda'\tilde{\delta}$ is a *blue* and thus (a) says $\Lambda'\widehat{\delta}$ and $\Lambda'\tilde{\delta}$ are *blues* for the same parametric vector and hence they are equal.□

Note that (c) gives the condition for $\mathrm{SSD}_d = \mathrm{SSD_R}$. Two obvious cases when this is true are the following:

- r$(X_dB) = $ p where p is length of Y_d (becasue then $\mathrm{R}(X_dB) = \mathbb{R}^{\mathrm{p}}$). An example here is the two-way model with interaction and no missing cells.
- $V = cI$ for some constant c (because $\mathrm{R}(L_{\mathrm{R}}) \subset \mathrm{R}(X_dB)$). An example here would be a balanced design, i.e., the same number of observations in each cell.

Note that if either of these conditions is true, then (c) is true for all estimable $\Lambda'\delta$. The following example illustrates a case where neither of the above conditions are true. That is, where (c) is true for some estimable parametric vectors, but not for all estimable parametric vectors.

Example 4.3. Consider the two-way additive model with $a = b = 3$. Using the constrained parametrization the model may be written as

$$\mathrm{E}(Y_{dij}) = \mu^c + \alpha_i^c + \beta_j^c\,, \quad i, j = 1,2,3\,,$$

where the parameters are constrained by $\Sigma_i \alpha_i^c = \Sigma_j \beta_j^c = 0$. Also, consider following artificial data

	B			
	Level	1	2	3
A	1	10	11	11
	2	15 14	19 16	14 15
	3	15 14 15 16 14 18	25 25 24 24 25 25	20 20 21 22 19 18

Note that this data set has the following n-pattern:

$$[n_{ij}] = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 6 & 6 & 6 \end{bmatrix}.$$

To compute $\widehat{\delta}$ and $\widetilde{\delta}$ write the model in matrix form as $\mathrm{E}(Y) = X\delta^c, \Delta'\delta^c = 0$. Then use Corollary 3.4 with $\mathrm{R}(B) = \mathrm{N}(\Delta')$. A satisfactory choice for B is

$$\Delta = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Using Corollary 3.4 this leads to

$\widehat{\delta}^c = [15.3889 \quad -4.7222 \quad 0.1111 \quad 4.6111 \quad -3.4074 \quad 3.5926 \quad -0.1852\,]'$ and

$\widetilde{\delta}^c = [15.3889 \quad -4.7222 \quad 0.1111 \quad 4.6111 \quad -2.1111 \quad 2.3333 \quad -0.2222\,]'$.

Suppose we want to test the row effects. This hypothesis can be written as

H: $\Lambda'\delta^c = 0$ where

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

In this case we can show that $\mathrm{R}(VL_{\mathrm{R}}) \subset \mathrm{R}(X_dB)$. That is,

$$\Lambda'\widehat{\delta} = \Lambda'\widetilde{\delta} = \begin{bmatrix} -4.8333 \\ -9.3333 \end{bmatrix}.$$

The basic summary statistics are:

$$\begin{array}{lll} \mathrm{SSD} = 270.7963 & \mathrm{SSD}_d = 130.7222 & \mathrm{SSD}_{\mathrm{R}} = 130.7222. \\ \mathrm{MSD} = 135.3981 & \mathrm{MSD}_d = 117.6500 & \mathrm{MSD}_{\mathrm{R}} = 117.6500. \end{array}$$

Now suppose we want to test for no column effects. This can be written as

H: $\Lambda'\delta^c = 0$ where

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

In this case $\mathrm{R}(VL_{\mathrm{R}})$ is not contained $\mathrm{R}(X_dB)$. The basic statistics are

$$\Lambda'\widehat{\delta} = \begin{bmatrix} -7.0000 \\ -3.2222 \end{bmatrix} \text{ and } \Lambda'\widetilde{\delta} = \begin{bmatrix} -4.4444 \\ -1.8889 \end{bmatrix},$$

and the summary statistics for the hypothesis tests are

$$\begin{array}{lll} \mathrm{SSD} = 220.9630 & \mathrm{SSD}_d = 73.6543 & \mathrm{SSD}_{\mathrm{R}} = 29.8519 \\ \mathrm{MSD} = 110.4815 & \mathrm{MSD}_d = 110.4815 & \mathrm{MSD}_{\mathrm{R}} = 26.8667. \ \square \end{array}$$

In terms of the distribution of $\mathcal{F}_{\mathrm{R}}$ let us make several comments. First, a theorem like Theorem 4.6 is available by using $\mathbb{C}_{\mathrm{R}}$ in place of $\mathbb{C}$ in the H matrix and in the noncentrality parameters. Also, a result like Corollary 4.7(a,c,d,e,f) is available by replacing $\mathbb{C}$ with $\mathbb{C}_{\mathrm{R}}$. Thus, an expression like (4.4) is available except that the denominator is not necessarily independent of the numerator sum of squares. Except for a couple of examples in the next chapter, we do not propose to further consider $\mathcal{F}_{\mathrm{R}}$ for the following reasons:

- $\mathcal{F}$ and $\mathcal{F}_d$ are functions of the minimal sufficient statistic, but this is not the case when $\Lambda'\widehat{\delta} \neq \Lambda'\widetilde{\delta}$.

- From our power calculations in the next chapter it seems that the test with the largest noncentrality parameter tends to have the best power. Thus, the fact that $\varphi_{\mathrm{R}} \leq \varphi_d$ would indicate that $\mathcal{F}_d$ has better power than $\mathcal{F}_{\mathrm{R}}$.
- From our limited power calculations, it appears that $\mathcal{F}_d$ is always a better test than $\mathcal{F}_{\mathrm{R}}$.

5. COMPUTING TAIL PROBABILITIES

There are special cases, as outlined in Proposition 4.12, where $\mathcal{F}_d$ has an F-distribution, but in general this is not the case. The denominator of the test statistic does have a chi-squared distribution that is independent of the numerator, but the numerator typically does not. However, as indicated in (4.4), the distribution of the numerator is like a linear combination of independent chi-squared random variables which allows us to apply some approximations considered in the literature.

In this chapter we introduce three approximation methods for determining the critical value K_d for the UM test statistic $\mathcal{F}_d$. The first approximation method is by Box (1954) which approximates the numerator mean square by a χ^2 distribution by equating the first two moments of the numerator mean square to the first two moments of a chi-squared distribution. The second method is due to Imhof (1961) who provided exact and approximate methods for computing tail probabilities. The last method is known as a saddle-point approximation. For this case we use Liberman (1994) who provides a compact formula for computing the tail probabilities. Further, the Imhof method and the Saddlepoint approximation can be used to determine K_R, the critical value for $\mathcal{F}_R$. Also, in addition to determining critical values, we show how the Imhof (1961) method can be used for power calculations.

5.1. Approximation by a Chi-squared Distribution.

This approximation method is typically called a Satterthwaite approximation. The basic idea is to approximate the distribution of a linear combination of independent chi-squared random variables, say Q, by $g\chi^2(h)$ where g, h are selected so that both distributions have the same first two moments. This idea seems to have been introduced by Welch (1937) when Q is a linear combination of two independent chi-squared squared random variables and generalized to an arbitrary linear combination by Satterthwaite (1941). The approximation was studied by Box (1954) to approximate the distribution of a quadratic form in normal random variables. Box (1954) also stated the following lemma.

Lemma (Box 1954). If Q and Q' are independent quadratic forms and if $g\chi^2(h)$ and $g'\chi^2(h')$ are Satterthwaite approximations for Q and Q' respectively, then the distribution of Q/Q' is approximately $kF(h, h')$ where $k = (gh)/(g'h')$.

This approximation has been used to compute the critical values for $\mathcal{F}_d$ in the two-way model with interaction by Gosslee and Lucas (1965) and for the one-way and two-way with interaction models by Rankin (1974). For $\mathcal{F}_d$ the approximation would be as follows: Using (4.4) with $\sigma^2 = 1$ we get that MSD_d under the null hypothesis is distributed approximately as $g\chi^2(h)$ where

$$g = \frac{\sum_{i=1}^{m} \mathrm{r}_i c_i^2}{\sum_{i=1}^{m} \mathrm{r}_i c_i} \quad \text{and} \quad h = \frac{\left(\sum_{i=1}^{m} \mathrm{r}_i c_i\right)^2}{\sum_{i=1}^{m} \mathrm{r}_i c_i^2}.$$

Then Box's above lemma with $Q = \mathrm{MSD}_d$ and $Q' = \mathrm{MSE} \sim \chi^2(f)/f$ implies that under the null hypothesis $\mathcal{F}_d$ is approximately distributed as $F(h, f)$. In computing these expressions note that $\Sigma_i \mathrm{r}_i c_i = 1$.

5.2. Imhof (1961) Exact and Approximation.

Imhof (1961) provided exact and approximate methods for computing the distribution of quadratic forms in normal variables. Let $U \sim \mathrm{N}(\mu, V)$. Then Q, a quadratic form in U, can always be expressed as $Q = \Sigma_1^m c_i \chi^2(\mathrm{r}_i, \lambda_i)$. In this expression the chi-squareds are independent and the c_i, unlike those in (4.4), can be positive or negative. Imhof writes the probability $\Pr(Q > x)$ as

$$\Pr(\Sigma_i c_i\chi^2(\mathrm{r}_i, \lambda_i) > x) = \tfrac{1}{2} + \tfrac{1}{\pi}\int_0^\infty \frac{\sin(\theta(u))}{up(u)}du, \tag{5.1}$$

where $\theta(u) = \frac{1}{2}\left[\Sigma_i \mathrm{r}_i \tan^{-1}(c_i u) + \lambda_i c_i u(1 + c_i^2 u^2)^{-1}\right] - \frac{1}{2}xu$

and $\quad p(u) = \prod_i (1 + c_i^2 u^2)^{\frac{1}{4}\mathrm{r}_i} \exp\left\{\frac{1}{2}\Sigma_i \lambda_i c_i^2 \Big/ (1 + c_i^2 u^2)\right\}.$

This expression is an exact probability statement. The approximation comes about because in numerical work the integration in (5.1) will be over a finite range $0 \leq u \leq T$. The degree of approximation obtained will depend on two sources of error: (i) the 'error of integration' and (ii) the 'error of truncation'. One can find that $|t|$, the error of truncation, can be bounded by T_u where

$$T_u^{-1} = \pi k T^k \prod_i |c_i|^{\frac{1}{2}\mathrm{r}_i} \exp\left\{\tfrac{1}{2}\Sigma_i \lambda_i c_i^2 T^2 (1 + c_i^2 T^2)^{-1}\right\}, \tag{5.2}$$

and where $k = \frac{1}{2}\Sigma_i \mathrm{r}_i$.

In statistical applications, it is sometimes possible to arrange for even degrees of freedom. In this case the following lemma provides an exact probability for a linear combination of independent central chi-squared random variables.

Lemma 5.1. Suppose $c_1, \cdots, c_m$ are ordered so that $c_1 > \ldots > c_p > 0 > c_{p+1} > \ldots > c_m$. Then for $x > 0$ we can write

$$\Pr\left[\Sigma_i^m c_i \chi^2(2r_i) > x\right] = \Sigma_k^p \frac{1}{(r_k - 1)!}\left[\frac{\partial^{r_k - 1}}{\partial \lambda^{r_k - 1}} F_k(\lambda, x)\right]_{\lambda = c_k},$$

where $F_k(\lambda, x) = \lambda^{n-1}\exp\{-x/(2\lambda)\}\prod_{i \neq k}^m (\lambda - c_i)^{-r_i}$ and $n = \Sigma_i^m \mathrm{r}_i$.

Lemma 5.1 has been explored by both Box (1954) and Imhof (1961). The form above is taken from Imhof.

Using (4.4), formula (5.1) can be easily applied to determining the power function for $\mathcal{F}_d$. Using (4.4) and its definitions given in Theorem 4.6 the power function for $\mathcal{F}_d$ can be expressed as

$$\beta_d = \Pr\{\mathcal{F}_d > K_d\} = \Pr\{\ \Sigma_1^s c_i \chi^2(\mathrm{r}_i, \lambda_i/\sigma^2) - K_d \chi^2(f)/f > 0\} \qquad (5.3)$$

where K_d is cutoff point satisfying

$$\Pr\{\ \Sigma_1^s c_i \chi^2(\mathrm{r}_i) - K_d \chi^2(f)/f > 0\} = \alpha.$$

Note that the above expressions for β_d fall exactly in line with (5.1) having $m = s + 1$ terms where the first s terms come from the numerator in (4.4), the remaining term is $-K_d\chi^2(f)/f$ and $x = 0$. To compute the cutoff point as in the next example we iterated using the above expression until achieving the desired α level. However, from a practical viewpoint it would be easier to compute the p-value and then judge the significance. That is, if $\mathcal{F}_d^*$ is the outcome for $\mathcal{F}_d$, then

$$\text{p-value} = \Pr\{\ \Sigma_1^s c_i \chi^2(\mathrm{r}_i) - \mathcal{F}_d^* \chi^2(f)/f > 0\}.$$

In using the Imhof formula (5.1) in the following examples the truncation error is always less than .00002. The integration was done using Version V Release 5.1 of Maple (1998). Thus, assuming the Maple integration is accurate, any power calculation done using the Imhof formula (5.1) in the following examples is essentially exact.

Example 5.1. Consider the two-way model with interaction in Example 4.4 where interest is in testing for no row effects. Assume that $a = b = 3$ and suppose we have the following two incidence patterns:

$$\text{Pattern 1} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 6 & 6 & 6 \end{bmatrix} \qquad \text{Pattern 2} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 7 & 5 & 5 \end{bmatrix}$$

The table below gives all of the relevant quantities in Example 4.4. Also, the power is calculated for $\theta = [\ 1\ 0\ 0\ ]'$ and $\theta = [\ 0\ 0\ 1\ ]'$. The power for $\mathcal{F}_d$ is given in the line identified as P_d and the power for $\mathcal{F}$ is in line P. The quantities corresponding to Example 4.4 are given in the columns labeled "Inter". The other columns labeled "Add" are for testing of no row effects in the additive model

$$\mathrm{E}(Y_{ijk}) = \mu + \alpha_i + \beta_j \,.$$

That is, the linear hypothesis for the additive model is that all of the α_i are equal. For this model the θ_i should be interpreted as

$$\theta_i = \mu + \alpha_i + \overline{\beta}_. \,.$$

Notice that the d_i are missing for the additive model. They could be calculated, but they do not define the c_i and λ_i as they do in the model with interaction

Table 5.1 Power comparison (Example 5.1)

θ'	[1 0 0]				[0 0 1]			
pattern	1		2		1		2	
model	Inter	Add	Inter	Add	Inter	Add	Inter	Add
d_1	1.0000		1.0000		1.0000		1.0000	
d_2	0.5000		0.5556		0.5000		0.5556	
d_3	0.1667		0.1810		0.1667		0.1810	
f	18	22	18	22	18	22	18	22
c_1	0.7179	0.7179	0.7045	0.7169	0.7179	0.7179	0.7045	0.7169
c_2	0.2821	0.2821	0.2955	0.2831	0.2821	0.2821	0.2955	0.2831
λ_1	2.4039	2.4039	2.3169	2.4882	0.2470	0.2470	0.1957	0.3123
λ_2	0.2627	0.2627	0.3228	0.1784	5.7530	5.7530	5.3795	5.6877
φ_d	1.8000	1.8000	1.7276	1.8342	1.8000	1.8000	1.7276	1.8342
φ	1.3333	1.3333	1.3198	1.3333	3.0000	3.0000	2.7876	2.9957
K_d	3.7083	3.5983	3.6888	3.5966	3.7083	3.5983	3.6888	3.5966
K	3.5546	3.4434	3.5546	3.4434	3.5546	3.4434	3.5546	3.4434
P_d	0.3026	0.3092	0.2943	0.3146	0.3031	0.3110	0.2942	0.3169
P	0.2508	0.2570	0.2483	0.2570	0.5097	0.5229	0.4791	0.5229

Notice that for the alternative $\theta' = (1\ 0\ 0)$ that $\mathcal{F}_d$ has more power than the usual test $\mathcal{F}$, but when $\theta' = (0\ 0\ 1)$ the reverse is true. That is, neither test is superior to the other. This observation was noted by Rankin (1974) in some simulations that he did in comparing $\mathcal{F}$ and $\mathcal{F}_d$ for testing for no treatment effects in the one-way model.□

Example 5.2. Let us again consider the situation in Example 4.4. In this case $s = 2$, $c_1 + c_2 = 1$ and $\mathrm{r}_1 = \mathrm{r}_2 = 1$. Instead of giving specific incidence patterns and values for the parameters as we did in the previous example, we have computed the cutoff points for various c_1 values and the power for the same c_1 values as well as various values of the noncentrality parameters and various f values. The tables are mostly self-explanatory. One particular thing you might note is that the $c_1 = .5$ column corresponds to the usual test statistic $\mathcal{F}$. The K_d line of each table gives the cutoff points and the values in the body of the tables give the power. The calculations are done using the Imhof formula.

Table 5.2 Power table of linear combination of chi-squared distribution

	$f = 5$						
c_1	0.1	0.2	0.4	0.5	0.6	0.8	0.9
K_d	6.2799	6.0449	5.8118	5.7861	5.8118	6.0449	6.2799
$\lambda_1 = 1$ $\lambda_2 = 0$	0.0550	0.0631	0.0847	0.0958	0.1061	0.1221	0.1271
$\lambda_1 = 2$ $\lambda_2 = 0$	0.0604	0.0777	0.1224	0.1448	0.1651	0.1961	0.2057
$\lambda_1 = 5$ $\lambda_2 = 0$	0.0791	0.1287	0.2458	0.2986	0.3439	0.4088	0.4284
$\lambda_1 = 1$ $\lambda_2 = 1$	0.1336	0.1378	0.1439	0.1448	0.1439	0.1378	0.1336
$\lambda_1 = 2$ $\lambda_2 = 1$	0.1405	0.1545	0.1837	0.1956	0.2047	0.2133	0.2132
$\lambda_1 = 5$ $\lambda_2 = 1$	0.1629	0.2095	0.3079	0.3491	0.3828	0.4261	0.4363
$\lambda_1 = 1$ $\lambda_2 = 2$	0.2132	0.2133	0.2047	0.1956	0.1837	0.1545	0.1405
$\lambda_1 = 2$ $\lambda_2 = 2$	0.2210	0.2311	0.2453	0.2472	0.2453	0.2311	0.2210
$\lambda_1 = 5$ $\lambda_2 = 2$	0.2454	0.2876	0.3677	0.3982	0.4210	0.4435	0.4444
$\lambda_1 = 1$ $\lambda_2 = 5$	0.4363	0.4261	0.3828	0.3491	0.3078	0.2095	0.1629
$\lambda_1 = 2$ $\lambda_2 = 5$	0.4444	0.4435	0.4210	0.3982	0.3677	0.2876	0.2454
$\lambda_1 = 5$ $\lambda_2 = 5$	0.4686	0.4952	0.5288	0.5331	0.5288	0.4952	0.4686

	$f = 10$						
c_1	0.1	0.2	0.4	0.5	0.6	0.8	0.9
K_d	4.6521	4.3983	4.1320	4.1028	4.1320	4.3983	4.6521
$\lambda_1 = 1$ $\lambda_2 = 0$	0.0545	0.0637	0.0945	0.1107	0.1247	0.1427	0.1465
$\lambda_1 = 2$ $\lambda_2 = 0$	0.0595	0.0799	0.1453	0.1770	0.2047	0.2387	0.2458
$\lambda_1 = 5$ $\lambda_2 = 0$	0.0777	0.1429	0.3168	0.3877	0.4418	0.5045	0.5176
$\lambda_1 = 1$ $\lambda_2 = 1$	0.1538	0.1614	0.1754	0.1776	0.1754	0.1614	0.1538
$\lambda_1 = 2$ $\lambda_2 = 1$	0.1614	0.1819	0.2299	0.2477	0.2585	0.2601	0.2547
$\lambda_1 = 5$ $\lambda_2 = 1$	0.1870	0.2531	0.4004	0.4542	0.4920	0.5259	0.5272
$\lambda_1 = 1$ $\lambda_2 = 2$	0.2547	0.2601	0.2585	0.2477	0.2299	0.1819	0.1614
$\lambda_1 = 2$ $\lambda_2 = 2$	0.2639	0.2827	0.3139	0.3183	0.3139	0.2827	0.2639
$\lambda_1 = 5$ $\lambda_2 = 2$	0.2935	0.3566	0.4780	0.5169	0.5404	0.5423	0.5368
$\lambda_1 = 1$ $\lambda_2 = 5$	0.5272	0.5259	0.4920	0.4542	0.4004	0.2531	0.1870
$\lambda_1 = 2$ $\lambda_2 = 5$	0.5362	0.5473	0.5404	0.5169	0.4780	0.3566	0.2935
$\lambda_1 = 5$ $\lambda_2 = 5$	0.5661	0.6108	0.6696	0.6770	0.6696	0.6108	0.5661

(Table 5.2 continued)

					$f = 27$			
c_1		0.1	0.2	0.4	0.5	0.6	0.8	0.9
K_d		3.9194	3.6654	3.3844	3.3541	3.3844	3.6654	3.9194
$\lambda_1 = 1$	$\lambda_2 = 0$	0.0542	0.0634	0.1027	0.1235	0.1407	0.1584	0.1607
$\lambda_1 = 2$	$\lambda_2 = 0$	0.0588	0.0802	0.1650	0.2057	0.2374	0.2702	0.2746
$\lambda_1 = 5$	$\lambda_2 = 0$	0.0757	0.1515	0.3768	0.4580	0.5140	0.5676	0.5750
$\lambda_1 = 1$	$\lambda_2 = 1$	0.1685	0.1790	0.2020	0.2057	0.2020	0.1790	0.1685
$\lambda_1 = 2$	$\lambda_2 = 1$	0.1767	0.2021	0.2687	0.2914	0.3025	0.2945	0.2845
$\lambda_1 = 5$	$\lambda_2 = 1$	0.2040	0.2856	0.4744	0.5340	0.5707	0.5913	0.5855
$\lambda_1 = 1$	$\lambda_2 = 2$	0.2845	0.2945	0.3025	0.2914	0.2687	0.2021	0.1767
$\lambda_1 = 2$	$\lambda_2 = 2$	0.2947	0.3206	0.3693	0.3764	0.3693	0.3206	0.2947
$\lambda_1 = 5$	$\lambda_2 = 2$	0.3276	0.4074	0.5614	0.6033	0.6240	0.6149	0.5961
$\lambda_1 = 1$	$\lambda_2 = 5$	0.5855	0.5913	0.5707	0.5340	0.4743	0.2856	0.2040
$\lambda_1 = 2$	$\lambda_2 = 5$	0.5961	0.6149	0.6240	0.6033	0.5614	0.4074	0.3276
$\lambda_1 = 5$	$\lambda_2 = 5$	0.6281	0.6846	0.7587	0.7674	0.7587	0.6846	0.6281
					$f = 50$			
c_1		0.1	0.2	0.4	0.5	0.6	0.8	0.9
K_d		3.7500	3.4981	3.2131	3.1826	3.2131	3.4981	3.7500
$\lambda_1 = 1$	$\lambda_2 = 0$	0.0541	0.0632	0.1053	0.1276	0.1453	0.1630	0.1649
$\lambda_1 = 2$	$\lambda_2 = 0$	0.0586	0.0800	0.1712	0.2145	0.2474	0.2792	0.1649
$\lambda_1 = 5$	$\lambda_2 = 0$	0.0750	0.1535	0.3953	0.4788	0.5344	0.5844	0.5903
$\lambda_1 = 1$	$\lambda_2 = 1$	0.1728	0.1840	0.2102	0.2145	0.2102	0.1840	0.1728
$\lambda_1 = 2$	$\lambda_2 = 1$	0.1812	0.2078	0.2807	0.3048	0.3158	0.3043	0.2930
$\lambda_1 = 5$	$\lambda_2 = 1$	0.2089	0.2948	0.4963	0.5569	0.5925	0.6086	0.6011
$\lambda_1 = 1$	$\lambda_2 = 2$	0.2930	0.3043	0.3158	0.3048	0.2807	0.2078	0.1812
$\lambda_1 = 2$	$\lambda_2 = 2$	0.3035	0.3313	0.3860	0.3940	0.3860	0.3313	0.3035
$\lambda_1 = 5$	$\lambda_2 = 2$	0.3373	0.4218	0.5852	0.6273	0.6468	0.6328	0.6119
$\lambda_1 = 1$	$\lambda_2 = 5$	0.6011	0.6086	0.5925	0.5569	0.4963	0.2948	0.2089
$\lambda_1 = 2$	$\lambda_2 = 5$	0.6119	0.6328	0.6468	0.6273	0.5852	0.4218	0.3373
$\lambda_1 = 5$	$\lambda_2 = 5$	0.6446	0.7039	0.7813	0.7901	0.7813	0.7039	0.6446

(Table 5.2 continued)

	$f = 100$						
c_1	0.1	0.2	0.4	0.5	0.6	0.8	0.9
K_d	3.6559	3.4050	3.1179	3.0873	3.1179	3.4050	3.6559
$\lambda_1 = 1$ $\lambda_2 = 0$	0.0540	0.0631	0.1069	0.1301	0.1483	0.1657	0.1673
$\lambda_1 = 2$ $\lambda_2 = 0$	0.0585	0.0798	0.1751	0.2199	0.2534	0.2845	0.2876
$\lambda_1 = 5$ $\lambda_2 = 0$	0.0746	0.1545	0.4066	0.4911	0.5465	0.5942	0.5992
$\lambda_1 = 1$ $\lambda_2 = 1$	0.1754	0.1870	0.2153	0.2199	0.2153	0.1870	0.1754
$\lambda_1 = 2$ $\lambda_2 = 1$	0.1838	0.2112	0.2881	0.3130	0.3289	0.3102	0.2980
$\lambda_1 = 5$ $\lambda_2 = 1$	0.2118	0.3003	0.5095	0.5704	0.6054	0.6187	0.6101
$\lambda_1 = 1$ $\lambda_2 = 2$	0.2980	0.3102	0.3289	0.3130	0.2881	0.2112	0.1838
$\lambda_1 = 2$ $\lambda_2 = 2$	0.3087	0.3377	0.3960	0.4046	0.3960	0.3377	0.3087
$\lambda_1 = 5$ $\lambda_2 = 2$	0.3430	0.4304	0.5994	0.6414	0.6601	0.6432	0.6210
$\lambda_1 = 1$ $\lambda_2 = 5$	0.6101	0.6187	0.6054	0.5704	0.5095	0.3003	0.2118
$\lambda_1 = 2$ $\lambda_2 = 5$	0.6210	0.6432	0.6601	0.6414	0.5994	0.4304	0.3430
$\lambda_1 = 5$ $\lambda_2 = 5$	0.6541	0.7150	0.7942	0.8030	0.7942	0.7150	0.6541
	$f = \infty$						
c_1	0.1	0.2	0.4	0.5	0.6	0.8	0.9
K_d	3.5649	3.3155	3.0265	2.9957	3.0265	3.3155	3.5649
$\lambda_1 = 1$ $\lambda_2 = 0$	0.0540	0.0630	0.1085	0.1327	0.1514	0.1685	0.1698
$\lambda_1 = 2$ $\lambda_2 = 0$	0.0584	0.0796	0.1791	0.2255	0.2597	0.2900	0.2926
$\lambda_1 = 5$ $\lambda_2 = 0$	0.0743	0.1555	0.4183	0.5038	0.5585	0.6039	0.6080
$\lambda_1 = 1$ $\lambda_2 = 1$	0.1780	0.1901	0.2205	0.2254	0.2205	0.1901	0.1780
$\lambda_1 = 2$ $\lambda_2 = 1$	0.1865	0.2146	0.2956	0.3215	0.3322	0.3161	0.3031
$\lambda_1 = 5$ $\lambda_2 = 1$	0.2149	0.3059	0.5229	0.5840	0.6182	0.6287	0.6190
$\lambda_1 = 1$ $\lambda_2 = 2$	0.3031	0.3161	0.3322	0.3215	0.2956	0.2146	0.1865
$\lambda_1 = 2$ $\lambda_2 = 2$	0.3140	0.3442	0.4063	0.4154	0.4063	0.3442	0.3140
$\lambda_1 = 5$ $\lambda_2 = 2$	0.3488	0.4391	0.6135	0.6554	0.6733	0.6535	0.6301
$\lambda_1 = 1$ $\lambda_2 = 5$	0.6190	0.6287	0.6182	0.5840	0.5229	0.3059	0.2149
$\lambda_1 = 2$ $\lambda_2 = 5$	0.6301	0.6535	0.6733	0.6554	0.6135	0.4391	0.3488
$\lambda_1 = 5$ $\lambda_2 = 5$	0.6635	0.7260	0.8066	0.8154	0.8066	0.7260	0.6635

There are a number of interesting observations that come from these tables. We mention just two of them. First, as might be expected, $\mathcal{F}_d$ has better power than $\mathcal{F}$ when the largest eigenvalue is associated with the largest noncentrality parameter. For example, compare $c_1 = .5$ (usual test) with $c_1 = .9$ when $\lambda_1 > \lambda_2$. Secondly, which is somewhat of a surprise, notice that in some situations the power of $\mathcal{F}_{\mathrm{d}}$ can decrease as f increases. For example, compare $c_1 = .1$, $\lambda_1 = 1$ and $\lambda_2 = 0$ as f increases.□

The previous two examples are computations for $\mathcal{F}_d$. One actually could use the Imhof formula (5.1) for $\mathcal{F}_{\mathrm{R}}$ even though the denominator and numerator are not necessarily independent. To see this first write $\mathcal{F}_{\mathrm{R}} = \frac{Y'AY}{Y'BY}$. Because B is a nonnegative definite matrix, note that we can write

$$\Pr(\, \mathcal{F}_{\mathrm{R}} > x) = \Pr(\, Y'DY > 0\,) \quad , \quad D = A - xB\,. \tag{5.4}$$

Now one can express $Y'DY$ in terms of a linear combination of independent chi-squareds and then apply the Imhof formula (5.1). To apply (5.1) in this context note that

$$B = f^{-1}(I - P_{\Omega})$$

and

$$A = (1/\mathrm{trace}(\mathbb{C}_{\mathrm{d}}^{+}\mathbb{C}_{\mathrm{R}}))H_{\mathrm{R}}\mathbb{C}_{d}^{+}H_{\mathrm{R}}'$$

where $H_{\mathrm{R}} = T(T'T)^{-1}L_{\mathrm{R}}$ with L_{R} defined as in Lemma 4.14. Although one could use this formulation to get $\Pr(\, \mathcal{F}_{\mathrm{R}} > x)$ we explore another method for getting $\Pr(\, \mathcal{F}_{\mathrm{R}} > x)$ under the null hypothesis in the following section.

5.3. Saddlepoint Approximation.

Lieberman (1994) suggested a saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables of the form $R = \frac{Z'AZ}{Z'BZ}$ where B is a nonnegative definite matrix and where $Z \sim \mathrm{MVN}(0,I)$. Liberman (1994) adopted the Lugannani and Rice (1980) formula which is applied in obtaining an explicit expression for the cumulative distribution function of the ratio of quadratic forms. To calculate tail probabilities, write one minus the cumulative generating function as

$$\Pr(R \geq x) = \Pr(Z'DZ \geq 0) = \tfrac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp\Big\{ -\tfrac{1}{2}|I - 2wD|\Big\} \tfrac{dw}{w}, \qquad (5.5)$$

where $D = A - xB$. Applied to (5.5), the Lugannani-Rice (1980) formula is

$$\Pr(R \geq x) \approx 1 - \Phi(\widehat{\xi}\,) + \phi(\widehat{\xi}\,)(\tfrac{1}{\widehat{z}} - \tfrac{1}{\widehat{\xi}}\,) \qquad (5.6)$$

where Φ and ϕ are the standard normal cdf and pdf respectively and where

$$\widehat{\xi} = sign(\widehat{\gamma})\sqrt{\Sigma_i \; ln(1 - 2\widehat{\gamma}d_i)} \quad ,$$

and

$$\widehat{z} = \widehat{\gamma}\sqrt{2\,\Sigma_i\{d_i^2/(1 - 2\widehat{\gamma}d_i)^2\}}.$$

In these expressions $\widehat{\gamma}$ is the saddlepoint satisfying $\Sigma_i \frac{d_i}{1-2\widehat{\gamma}d_i} = 0$ and $d_1, \ldots, d_n$ are the eigenvalues of the $n \times n$ matrix D.

Now let us apply the above to $\mathcal{F}_{\mathrm{R}}$. First note that we may suppose $\sigma^2 = 1$. Next, note that we can take A and B as defined at the end of the previous section. Now let us examine the distribution under the null hypothesis. Under the null hypothesis H: $\Lambda'\delta = 0$ we may suppose that Y has the same distribution as $Z + \mu$ where $\mu \in \Omega_{\mathrm{H}}$. However, it can be shown that $A\mu = 0$ and $B\mu = 0$ so that under the null hypothesis the distribution of $\mathcal{F}_{\mathrm{R}}$ fits the above assumptions for the saddlepoint approximation.

To get a critical value for $\mathcal{F}_{\mathrm{R}}$ notice that one must iterate. However, as with the Imhof calculations, it would be easier to simply use (5.6) to compute a p-value with $x = \mathcal{F}_{\mathrm{R}}^*$ where $\mathcal{F}_{\mathrm{R}}^*$ is an outcome on $\mathcal{F}_{\mathrm{R}}$ and then judge the significance based on the p-value.

5.4. Critical Values.

To compute the critical values for $\mathcal{F}_d$ one can use any of the three methods previously discussed. For $\mathcal{F}_\mathrm{R}$, however, only the Imhof (1964) and the Saddlepoint approximation are appropriate when $\mathcal{F}_\mathrm{R} \neq \mathcal{F}_d$, i.e., when the numerator and the denominator of $\mathcal{F}_\mathrm{R}$ are not independent.

To get an idea about the three approximations several examples were examined. Two tables are given below. The first table is for a two-way model with interaction and the linear hypothesis is that of no row effects. The second table is for the two-way additive model and again the linear hypothesis is that of no row effects. Each table gives a specific incidence pattern labeled the n-pattern. The r and f columns in the table are defined in the usual way and the h column is the approximate degrees of freedom for the chi-squared approximation for SSD_d, e.g., see the last part of Section 5.1. Then for each of the three possible test statistics $\mathcal{F}$, $\mathcal{F}_\mathrm{R}$ and $\mathcal{F}_d$ identified via the rows of the table the critical points for the three methods are given in the last three columns of the table. Note that in the first table the $\mathcal{F}_\mathrm{R}$ row is not included because $\mathcal{F}_\mathrm{R} = \mathcal{F}_d$ is always true in the two-way model with interaction. Also, in the second table the chi-squared approximation is included for $\mathcal{F}_\mathrm{R}$ even though it is not generally appropriate because the numerator and the denominator are not independent.

Table 5.3 Critical value (Testing Row effect, $\alpha = 0.05$) : Interaction Model

	n-pattern	r	h	f	F-dist	Imhof	Saddle Point
$\mathcal{F}$	1 1 1	2.0000	2.0000	18.0000	3.5546	3.5546	3.5751
$\mathcal{F}_d$	2 2 2	2.0000	1.6807	18.0000	3.7490	3.7083	3.7001
	6 6 6						
$\mathcal{F}$	1 2 6	2.0000	2.0000	18.0000	3.5546	3.5546	3.5751
$\mathcal{F}_d$	1 2 6	2.0000	2.0000	18.0000	3.5546	3.5546	3.5751
	1 2 6						
$\mathcal{F}$	1 1 1	2.0000	2.0000	18.0000	3.5546	3.5546	3.5751
$\mathcal{F}_d$	1 3 3	2.0000	1.7134	18.0000	3.7267	3.7083	3.6748
	7 5 5						
$\mathcal{F}$	1 2 6	2.0000	2.0000	18.0000	3.5546	3.5546	3.5751
$\mathcal{F}_d$	6 1 2	2.0000	2.0000	18.0000	3.5546	3.5546	3.5751
	2 6 1						

Table 5.4 Critical value (Testing Row effect, $\alpha = 0.05$) : Additive Model

	n-pattern	r	h	f	F-dist	Imhof	Saddle Point
$\mathcal{F}$	1 1 1	2.0000	2.0000	22.0000	3.4434	3.4434	3.4611
$\mathcal{F}_d$	2 2 2	2.0000	1.6807	22.0000	3.6374	3.5983	3.5900
$\mathcal{F}_{\mathrm{R}}$	6 6 6	2.0000	1.6807	22.0000	3.6374	3.5983	3.5900
$\mathcal{F}$	1 2 6	2.0000	2.0000	22.0000	3.4434	3.4434	3.4611
$\mathcal{F}_d$	1 2 6	2.0000	2.0000	22.0000	3.4434	3.4434	3.4611
$\mathcal{F}_{\mathrm{R}}$	1 2 6	2.0000	2.0000	22.0000	3.4434	3.2080[(1)]	3.2237
$\mathcal{F}$	1 1 1	2.0000	2.0000	22.0000	3.4434	3.4434	3.4611
$\mathcal{F}_d$	1 3 3	2.0000	1.6833	22.0000	3.6356	3.5968	3.5879
$\mathcal{F}_{\mathrm{R}}$	7 5 5	2.0000	1.7134	22.0000	3.6151	3.5459[(2)]	3.5344
$\mathcal{F}$	1 2 3	2.0000	2.0000	22.0000	3.4434	3.4434	3.4611
$\mathcal{F}_d$	1 2 0	2.0000	1.8243	22.0000	3.5441	3.5195	3.4889
$\mathcal{F}_{\mathrm{R}}$	7 5 6	2.0000	1.7672	22.0000	3.5799	3.4787[(3)]	3.4562
$\mathcal{F}$	0 1 2	2.0000	2.0000	3.0000	9.5521	9.5221	9.7590
$\mathcal{F}_d$	1 0 1	2.0000	1.9836	3.0000	9.5583	9.5553	9.7153
$\mathcal{F}_{\mathrm{R}}$	2 1 0	2.0000	1.9802	3.0000	9.5596	9.4825[(4)]	9.6395

Note that in Table 5.4, four of the Imhof critical values have superscripts. These critical values are where the numerator and the denominator of $\mathcal{F}_{\mathrm{R}}$ are not independent. When the numerator and denominator are independent as in $\mathcal{F}_d$, the Imhof calculations in the above table are based on three independent chi-squares with 1, 1, and f degrees of freedom. However, when they are not independent, the number of independent chi-squareds that are needed to do calculations is changed. To see how the non-independence affects that number of chi-squareds we have given below the calculations corresponding to each of the critical values with a superscript. These are:

(1) $\Pr\{0.4511\chi^2(2) - 0.0970\chi^2(2) - 0.1458\chi^2(20) > 0\} = 0.05$

(2) $\Pr\{0.6987\chi^2(1) + 0.2857\chi^2(1) - 0.1455\chi^2(1) - 0.1611\chi^2(1) - 0.1612\chi^2(20) > 0\} = 0.05$

(3) $\Pr\{0.6607\chi^2(1) + 0.3134\chi^2(1) - 0.1361\chi^2(1) - 0.1542\chi^2(1) + 0.1581\chi^2(20) > 0\} = 0.05$

(4) $\Pr\{0.5415\chi^2(1) + 0.4500\chi^2(1) - 3.1523\chi^2(1) - 3.1608\chi^2(2) > 0\} = 0.05$

In these calculations note that the sum of degrees of freedom corresponding to negative coefficients is the error degrees of freedom f and the sum of the degrees of freedom corresponding to the positive coefficients is equal to $r = 2$.

In these tables the Imhof critical values are essentially exact. Thus, it can be observed that the chi-square approximation is pretty good for $\mathcal{F}_{\mathrm{d}}$, but it not very good for $\mathcal{F}_{\mathrm{R}}$ in designs where the numerator and denominator are dependent. Also, except for the last n-pattern in the second table the Saddlepoint approximation seems satisfactory.

5.5. Some Power Comparisons.

Some power comparisons for $\mathcal{F}$ and $\mathcal{F}_d$ were done in Section 5.2. In this section we give a small power comparison study which includes $\mathcal{F}_{\mathrm{R}}$. The table below gives the comparisons The model is the two-way additive model

$$\mathrm{E}(Y_{ijk}) = \mu + \alpha_i + \beta_j,$$

and the linear hypothesis is that of no row effects. That is, the linear hypothesis is that all of the α_i are equal. The table gives two different alternatives. The first alternative has $\alpha_1 = 1$ and all other parameters zero where the second alternative has $\alpha_3 = 1$ and all other parameters zero. The distributions of the three mean squares MSD, MSD_d and $\mathrm{MSD}_{\mathrm{R}}$ have the form

$$c_1\chi^2(1,\lambda_1/\sigma^2) + c_2\chi^2(1,\lambda_2/\sigma^2)$$

where the c_i and λ_i are given in the line corresponding to the mean square for the three possible mean squares. Also, the three possible "noncentrality parameters" φ, φ_d and φ_{R} are given in the φ column. The power calculations are given in the last column. All power calculations for $\mathcal{F}$ and $\mathcal{F}_d$ are computed via the Imhof (5.1) and (5.3) formulae and the power calculations for $\mathcal{F}_{\mathrm{R}}$ are computed via (5.1) and (5.4) except in the in the second n-pattern where $\mathcal{F}_d = \mathcal{F}_{\mathrm{R}}$. For example, the calculations for the 3rd n-pattern for the $\alpha_1 = 1$ alternatives are

$$\Pr\{\mathcal{F} > K\} = \Pr\{0.5\chi^2(1,2.4882) - 0.5\chi^2(1,0.1784) - 0.1565\chi^2(22) > 0\} = .257$$

$$\Pr\{\mathcal{F}_d > K_d\} = \Pr\{0.7169\chi^2(1,2.4882) - 0.2831\chi^2(1,0.1784) - 0.1635\chi^2(22) > 0\} = .3146$$

$$\Pr\{\mathcal{F}_{\mathrm{R}} > K_{\mathrm{R}}\} = \Pr\{0.6987\chi^2(1,2.3664) + 0.2857\chi^2(1,0.2736) - 0.1455\chi^2(1,0.0258) - 0.1611\chi^2(1) - 0.1612\chi^2(20) > 0\} = .3044$$

Note that for $\mathcal{F}$ and $\mathcal{F}_d$

the linear combinations for computing the power can be determined from the table, but that is not true for $\mathcal{F}_{\mathrm{R}}$, except in the second n-pattern, because of the lack of independence between the numerator and the denominator.

Table 5.5 Power comparison (Additive Model)

	n-pattern		c_1	c_2	λ_1	λ_2	φ	power
$\alpha_1 = 1$	1 2 6	MSD	0.5000	0.5000	3.0000	3.0000	3.0000	0.5229
	1 2 6	MSD_d	0.5000	0.5000	3.0000	3.0000	3.0000	0.5229
	1 2 6	$\mathrm{MSD}_{\mathrm{R}}$	0.5000	0.5000	1.8000	1.8000	1.8000	0.3711
	1 1 1	MSD	0.5000	0.5000	2.4039	0.2627	1.3333	0.2570
	2 2 2	MSD_d	0.7179	0.2821	2.4039	0.2627	1.8000	0.3092
	6 6 6	$\mathrm{MSD}_{\mathrm{R}}$	0.7179	0.2821	2.4039	0.2627	1.8000	0.3092
	1 1 1	MSD	0.5000	0.5000	2.4882	0.1784	1.3333	0.2570
	1 3 3	MSD_d	0.7169	0.2831	2.4882	0.1784	1.8342	0.3146
	7 5 5	$\mathrm{MSD}_{\mathrm{R}}$	0.7045	0.2955	2.3169	0.3228	1.7276	0.3044
$\alpha_3 = 1$	1 2 6	MSD	0.5000	0.5000	0.4019	5.5981	3.0000	0.5229
	1 2 6	MSD_d	0.5000	0.5000	0.4019	5.5981	3.0000	0.5229
	1 2 6	$\mathrm{MSD}_{\mathrm{R}}$	0.5000	0.5000	0.2412	3.3588	1.8000	0.3711
	1 1 1	MSD	0.5000	0.5000	0.2470	5.7530	3.0000	0.5229
	2 2 2	MSD_d	0.7179	0.2821	0.2470	5.7530	1.8000	0.3110
	6 6 6	$\mathrm{MSD}_{\mathrm{R}}$	0.7179	0.2821	0.2470	5.7530	1.8000	0.3110
	1 1 1	MSD	0.5000	0.5000	0.3123	5.6877	3.0000	0.5229
	1 3 3	MSD_d	0.7169	0.2831	0.3123	5.6877	1.8342	0.3169
	7 5 5	$\mathrm{MSD}_{\mathrm{R}}$	0.7045	0.2955	0.1957	5.3795	1.7276	0.3064

There are two things in this table that are worth noting. First the ordering of the power is exactly correlated with the ordering of the noncentrality parameters. Secondly, the power of $\mathcal{F}_{\mathrm{d}}$ always exceeds the power of $\mathcal{F}_{\mathrm{R}}$. These two observations have also been observed in other examples that we have run.

6. CONCLUSIONS, CONJECTURES AND REMARKS

The unweighted means procedure was introduced in Yates (1934). Since that time a number of writers have suggested using the procedure in variance component estimation and confidence intervals in mixed linear models and in hypothesis testing in fixed linear models. The literature on variance component analysis is more extensive than the literature on hypothesis testing in fixed linear models. The purpose of this thesis has been to provide a more through and study of the unweighted means procedure in testing linear hypotheses in fixed linear models.

6.1. Conclusions.

Assume one has the linear model $\mathrm{E}(Y) = X\delta$, $\Delta'\delta = 0$, where $\mathrm{Cov}(Y) = \sigma^2 I$ and suppose one is interesting in testing a linear hypothesis H: $\theta = 0$ where $\theta = \Lambda'\delta$ is an estimable linear parametric vector. Let $\widehat{\theta}$ be the *blue* for θ and let $\mathbb{C}$ be such that $\mathrm{Cov}(\widehat{\theta}) = \sigma^2\mathbb{C}$. A summary of the results and conclusions in this thesis are briefly summarized below.

- **Construction of $\mathcal{F}_d$.** Traditionally the literature on testing linear hypotheses in the fixed linear model using the UM procedure has focused on the testing of main effects in the two-way additive model with interaction. In Section 2 of Chapter 4 we generalize these results to show how the UM test statistic $\mathcal{F}_d$ can be constructed for any linear model of the form specified above and for any linear hypothesis of the form H: $\theta = 0$.
- **The submodel.** The test statistic $\mathcal{F}_d$ depends on the investigator's choice of a submodel $\mathrm{E}(Y_d) = X_d\delta$, $\Delta'\delta = 0$. To get the submodel one writes $X = TX_d$ where T is $n \times d$ of rank d. Then $Y_d = (T'T)^{-1}T'Y$. Here the choice of T

and X_d are not unique. We have suggested choosing the rows of X_d as the distinct rows of X and then choosing T as the obvious classification matrix satisfying $X = TX_d$. That is, row i of T is all zeros except a one in the kth column where row k of X_d is the ith row of X. We think this partition of X is the one that most closely follows the spirit of the unweighted means analysis. There are, however, other choices. For example, consider a 2×2 additive model with the incidence pattern

$$[n_{ij}] = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

Then two possible T and X_d matrices are

$$T_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad X_{d1} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

and $$T_2 = \begin{pmatrix} 2 & -1 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad X_{d2} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The first T, X_d combination is what we would suggest, but the second also satisfies the requirements set out in Section 2 of Chapter 4. We did not pursue examining the differences in $\mathcal{F}_d$ resulting from different submodels.

- **The form of the test.** The form of the devSS for the usual test is $\text{SSD} = \widehat{\theta}'\mathbb{C}^{+}\widehat{\theta}$. For the UM test the devSS has the same form except that $\mathbb{C}_d$ is substituted for $\mathbb{C}$ where $\mathbb{C}_d$ is the covariance matrix of $\widehat{\theta}$ computed from the submodel acting as though $\text{Cov}(Y_d) = \sigma^2 I$ instead of the true $\text{Cov}(Y_d) = \sigma^2(T'T)^{-1}$. That is, the test statistic tries to mimic the test statistic that would be computed from the submodel. Typically the submodel is balanced so that the test statistic tries to mimic a balanced model. The hope is that this will tend to make the power function behave more like that of a balanced design.

- **Invariance of the test.** The devSS for the UM test is like the devSS for the usual test in the sense that no matter how the linear hypothesis is formulated nor computed the answer is always the same.

- **Distribution of the test.** The distribution of the test statistic $\mathcal{F}_d$ for the UM procedure has the form

$$\mathcal{F}_d \sim \frac{\Sigma_i c_i \chi^2(\mathrm{r}_i, \lambda_i/\sigma^2)}{\chi^2(f)/f} \tag{6.1}$$

where the c_i, r_i and λ_i are determined from the eigenvalues, multiplicities and eigenvectors of $\mathbb{C}^{\frac{1}{2}}\mathbb{C}_d^{+}\mathbb{C}^{\frac{1}{2}}$ as described in Theorem 4.6. Under the null hypothesis the λ_i are zero and under the alternative hypothesis the power of the test is an increasing function in each λ_i.

- **Critical values and power.** We showed how the Imhof (1961) results can be applied to basically get exact critical values, p-values and power. We also investigated the Satterthwaite approximation and the Saddlepoint approximation for getting critical values.

- **An alternative.** We briefly investigated in Section 4 of Chapter 6 the possibility of using the *blue* for θ computed from the submodel acting like the covariance matrix is $\sigma^2 I$ in the devSS for the UM method. However, from some simulations not included in the thesis and from our power computations we believe using the actual *blue* results in a better test.

- **Some observations.** The choice between the usual test and the UM procedure test is not easy because of fact that neither test is uniformly more powerful than the other. When the largest eigenvalue is associated the largest individual noncentrality parameters, then UM procedure seems more powerful. Otherwise the usual test is as good or better than UM procedure. From empirical evidence it seems that the noncentrality parameters φ and φ_d correlate with the power of the test. That is, when $\varphi > \varphi_d$ it seems that the usual test has the best power and the UM test based on $\mathcal{F}_d$ has the best power when $\varphi < \varphi_d$. Also, as stated in the literature the power of the usual test is very dependent on the configuration of the sample sizes and while the configuration does affect the power of $\mathcal{F}_d$ the effects on the power seem somewhat less when compared with the usual test $\mathcal{F}$. Finally, from the evidence we have seen in this thesis along with the ease of using $\mathcal{F}$, our overall conclusion is that generally the usual test would be preferred over the UM test. However, if one wanted to go to the work of getting the quantities in (6.1) and had an idea about the alternatives of interest, then it might be beneficial to use the Imhof procedure to calculate some power comparisons before deciding whether or not to use the usual test.

6.2. Remarks (probability inequalities).

In our study to determine the critical values K and K_d, we needed some probability inequalities. Although we were not able to prove exactly what we wanted we decided to include here some of the results that we did obtain as well as some conjectures. These probability inequalities are all concerned with nonnegative independent linear combinations of chi-squared random variables. Proofs for the lemmas below and an asymptotic proof of conjectures are in the appendix.

Let X_1 and X_2 be independent random variables with continuous density functions f_1,f_2 and distribution functions F_1,F_2 respectively. For $0 < \lambda < 1$, let

$T_\lambda = \lambda X_1 + (1-\lambda)X_2$ and let $G(t;\lambda) = \mathrm{P}\{T_\lambda \le t\}$ denote the distribution function of T_λ. The first two lemmas below were suggested by Professor Birkes.

Lemma 6.1. $G(t;\lambda) = \int_{-\infty}^{\infty} F_1(\lambda^{-1}(t-x)+x) f_2(x)\mathrm{d}x$.

The proof of the next lemma requires that one can differentiate through an integral. We did not rigorously verify that this can be done, but the result of the lemma is not used in any other results.

Lemma 6.2. $(\partial/\partial\lambda)G(t;\lambda) = -\lambda^{-2}\int_{-\infty}^{t/(1-\lambda)}(t-x)f_1(\lambda^{-1}(t-x)+x)f_2(x)\mathrm{d}x$.

Now for $\lambda_1,\ldots,\lambda_r \ge 0$ and $\Sigma_i\lambda_i = 1$, let

$$S_r = \Sigma_i\lambda_i X_i \quad \text{and} \quad g = \Sigma_i\lambda_i^2 ,$$

where $X_1, \cdots X_r$ are iid $\chi^2(1)$. Note that, MSD_d is of the form S_r whereas the usual MSD is distributed as $\frac{1}{r}\chi^2(r)$.

Conjecture 6.1. Suppose that $t > 0$. Then

(a) For sufficiently large t, $\Pr\{S_r \le t\}$ is a decreasing function of g.

(b) For sufficiently small t, $\Pr\{S_r \le t\}$ is an increasing function g.

By noting that the minimum value for g is $1/r$, that this minimum occurs at $\lambda_i = 1/r$ for $i = 1,\ldots,r$, and that $S_r \sim \frac{1}{r}\chi^2(r)$ when g attains this minimum we immediately get the following conjecture.

Conjecture 6.2.

(a) For sufficiently large t, $\Pr\{S_r \le t\} \le \mathrm{P}\left[\frac{1}{r}\chi^2(r) \le t\right]$.

(b) For sufficiently small t, $\Pr\{S_r \le t\} \ge \mathrm{P}\left[\frac{1}{r}\chi^2(r) \le t\right]$.

Note that Conjecture 6.2 provides a way to get inequalities for the critical points of $\mathcal{F}$ and $\mathcal{F}_d$. That is, for small α we would expect that Conjecture 6.2(a) would be applicable to give

$$\Pr\{\mathcal{F}_d > t\} \ge \Pr\{\mathcal{F} > t\} .$$

Which implies for small α that if $\Pr\{\mathcal{F}_d > K_d\} = \Pr\{\mathcal{F} > K\ \} = \alpha$, then $K_d > K$ which is consistent with our numerical calculations.

We cannot prove the above conjectures, but we have verified the conjectures for certain special cases. Our numerical work also suggests that the conjectures are true. In addition, using a Satterthwaite approximation for S_r and then a Wilson-Hilferty (1931) approximation we can verify the following lemma.

Lemma 6.3. For $t \geq 0$ and g as defined above, let

$$b(t,g) = \left(t^{1/3} - (1 - \tfrac{2}{9}g)\right) \div \sqrt{\tfrac{2}{9}\,g}$$

Then

(a) $\Pr\{S_r \leq t\} \doteq \Pr\{Z \leq b(t,g)\}$.

(b) $\frac{\partial}{\partial g} b(t,g)$ is positive if $t < (\frac{2g}{9} + 1)^3$ and negative if $t > (\frac{2g}{9} + 1)^3$.

(c) If $t > (\frac{2}{9} + 1)^3$ is fixed, then $\Pr\{Z \leq b(t,g)\}$ is a decreasing function of g.

(d) If $t < (\frac{2}{9r} + 1)^3$ is fixed, then $\Pr\{Z \leq b(t,g)\}$ is an increasing function of g.

Note that (a), (c) and (d) of this lemma provide an approximate justification for Conjectures 6.1, 6.2. In fact, parts (c) and (d) of the lemma provide guidelines for the "sufficiently large" and "sufficiently small" statements in Conjectures 6.1, 6.2.

BIBLIOGRAPHY

Anderson, R. L., and Bancroft, T. A. (1952), *Statistical Theory in Research*, New York, McGraw-Hill Book Co.

Box G. E. P. (1954), "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *Ann. Math. Statist.* 25, 290-302

Burdick, R. K. and Graybill, F. A. (1992), *Confidence Intervals on Variance Components,* New York, Marcel Dekker, Inc.

El-Bassiouni, M. Y. and El-Shahat M. A. T. (1986), "On testing treatment effects in the unbalanced random one-way model" *The Egyptian Statistical Journal*, 30, 65-91

Eubank L. E., Seely, J. F. and Lee, Y. (2001) Generalized Unweighted Means for the General Two Variance Components Mixed Model, , Unpublished manuscript. Department of Statistics, Oregon State University, Corvallis,

Francis I. (1973) "A Comparison of Several Analysis of variance Programs" *Journal of American Statistical Association*, 68, 860-865

Gosslee, D. G. and Lucas, H. L. (1965) "*Analysis of Variance of Disproportionate Data When Interaction is present,*' Biometrics, 21, 115-133

Henderson, C. R. (1978) "Variance-Corvariance matrix of Estimators of variances in Unweighted means ANOVA," *Biometrics*, 34, 462-468.

Imhof J. P. (1961) "Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika*, 48, 419-426

Knoke, J. D. (1985), "The Method of Unweighted Means in Unbalanced Analysis of Variance," *Biometrical Journal*, 27, 645-652.

Kutner, M. H. (1974) "Hypothesis Testing in Linear Model (Eisenhart Model I)", *The American Statistician*, 28, 98-100

Lehmann, E. L. (1983), *Theory of Point Estimation*, John Wiley & Sons

Levy, K. J., Narula, S. C, and Alrami, P. F. (1975), "An Empirical Comparison of the Methods of Least Squares and Unweighted means for the Analysis of Disproportionate Cell Data," *International Statistical review*, 43, 335-338.

Lieberman, Offer (1994), "Saddlepoint Approximation for the Distribution of a Ratio of Quadratic Forms in Normal Variables," *Journal of the American Statistical Association*, 89, 924-928.

Lugannani, R. and Rice, S. (1980) "Saddlepoint Approximation for the Distribution of the Sum of Independent Random Variables," *Advanced in Applied Probability*, 12, 475-490

Purdy, K. G., Seely, J. F. and Lee, Y. (1998) Confidence Intervals for variance Components, Unpublished manuscript. Department of Statistics, Oregon State University, Corvallis, Oregon

Rankin, N. O. (1974), "The Harmonic Mean Method for One-way and Two-way Analysis of Variance," *Biometrika*, 61, 117-122.

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

Senedecor, G. W. and Cochran, W. G. (1967) *Statistical methods*, Iowa State University Press, Ames, Iowa

Speed, F.M., Hocking, R. R., and Hackney, O. P. (1978) "Methods of Analysis of Linear Models with Unbalanced Data," *Journal of American Statistical Association*, 73, 105-112.

Speed, F. M., and Monlezun, C. J. (1979), "Exact F test for the Method of Unweighted Means in a 2k Experiment," *The American Statistician*, 33, 15-18

Thomas, J. D. and Hultquist, R. A. (1978), "Interval Estimation for the Unbalanced Case of the One-way Random Effect Model," The Annals of Statistics, 6, 582-587

Wilson, E. B and Hilferty (1931), "The Distribution of Chi-square", *Proceedings of National Academy of Science,* 17, 684-688

Yates, F. (1934), "The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes," *Journal of American Statistical Association*, 29, 51-66.

APPENDIX

APPENDIX

Proof of Lemma 6.1. Write

$G(t;\lambda) = \mathrm{P}\{T_\lambda \le t\} = \mathrm{E}[\mathrm{P}\{T_\lambda \le t \mid X_2\}] = \int_{-\infty}^{\infty} \mathrm{P}\{T_\lambda \le t \mid X_2 = x\} f_2(x)\mathrm{d}x$.

Because of independence we can write

$\mathrm{P}\{T_\lambda \le t \mid X_2 = x\} = \mathrm{P}\{\lambda X_1 + (1-\lambda)x \le t \mid X_2 = x\}$

$= \mathrm{P}\{\lambda X_1 + (1-\lambda)x \le t\} = F_1(\lambda^{-1}[t-(1-\lambda)x]) = F_1(\lambda^{-1}(t-x)+x).\square$

Proof of Lemma 6.2. By Lemma 1,

$$(\partial/\partial\lambda)G(t;\lambda) = (\partial/\partial\lambda)\int_{-\infty}^{\infty} F_1(\lambda^{-1}(t-x)+x)f_2(x)\mathrm{d}x$$

Note that $F_2(x)$ is not differentiable at $x = 0$. However, F_2 is differentiable almost everywhere. Now suppose that $G(t;\lambda)$ is differentiable and that the derivative could be taken through the integral sign. Then the argument would proceed as follows.

$(\partial/\partial\lambda)\int_{-\infty}^{\infty} F_1(\lambda^{-1}(t-x)+x)f_2(x)\mathrm{d}x = \int_{-\infty}^{\infty}(\partial/\partial\lambda)F_1(\lambda^{-1}(t-x)+x)f_2(x)\mathrm{d}x$. As x varies from 0 to ∞, $u = \lambda^{-1}(t-x)+x = \lambda^{-1}[t-(1-\lambda)x]$ varies from $\lambda^{-1}t$ to $-\infty$, but $F_1(u) = 0$ for $u \le 0$, i.e., for $x \ge t/(1-\lambda)$. Then $(\partial/\partial\lambda)G(t;\lambda) = \int_{-\infty}^{t/(1-\lambda)}(\partial/\partial\lambda)F_1(\lambda^{-1}(t-x)+x)f_2(x)\mathrm{d}x =$ $\int_{-\infty}^{t/(1-\lambda)}(-\lambda^{-2})(t-x)f_1(\lambda^{-1}(t-x)+x)f_2(x)\mathrm{d}x$. $\square$

Proof of Lemma 6.3.

First let us apply a Satterthwaite approximation to S_r. That is approximate S_r by $g\chi^2(h)$ where g and h are selected to have the same first two moments as S_r. This gives

$$g = \Sigma_i \lambda_i^2 \quad \text{and} \quad h = 1/\Sigma_i \lambda_i^2$$

Now employ the Wilson-Hilferty (1931) approximation which states that $(\chi^2(h)/h)^{1/3}$ is approximately normal with mean $(1 - \frac{2}{9h})$ and variance $\frac{2}{9h}$. Thus we may write

$$\Pr(S_r \leq t) \doteq \Pr\left\{\left(1 - \tfrac{2}{9h} + Z\sqrt{2/9h}\right)^3 \leq t\right\}$$

where Z is normally distributed with zero mean and unit variance. Thus, we can write

$$\Pr(S_r \leq t) \doteq \Pr\left\{Z \leq b(t, g)\right\}.$$

This gives part (a). For part (b) note that

$$\tfrac{\partial}{\partial g} b(t, d) = \tfrac{1}{12} \tfrac{\sqrt{2}(2g - 9t^{1/3} + 9)}{g^{3/2}},$$

from which part (b) follows. To get parts (c) and (d) use part (b) and the fact that $1/r \leq g \leq 1$.□