71-23,077

JOHNSON, Dallas Eugene, 1938-
INTERACTION IN TWO-WAY CLASSIFICATIONS WITHOUT REPLICATION.

Colorado State University, Ph.D., 1970
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THESIS

INTERACTION IN TWO-WAY CLASSIFICATIONS WITHOUT REPLICATION

Submitted by

Dallas Eugene Johnson

In partial fulfillment of the requirements
for the Degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
December, 1970

COLORADO STATE UNIVERSITY

December 1970

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY Dallas E. Johnson

ENTITLED Interaction in the Two-way Classifications Without Replication

BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY .

Committee on Graduate Work

Robert E. Gaines

Duane C. Boes

Paul W. Milke, Jr.

Franklin A. Graybill

Adviser

Franklin A. Graybill

Head of Department

ABSTRACT OF DISSERTATION

INTERACTIONS IN TWO-WAY CLASSIFICATIONS WITHOUT REPLICATION

Consider the two-way classification model with one observation per cell,

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + e_{ij} \quad i = 1, 2, \cdots, t; \; j = 1, 2, \cdots, b$$

with the usual identifiability and normality assumptions. When non-additivity in the model is demonstrated to be present, the procedure has been to find a transformation which will make the transformed data additive.

It seems that there is merit in performing an analysis of the data in the original units, if only for ease of interpretation. The procedure recommended in this thesis is to examine the data to find the pattern of interaction and to obtain a usable estimate of the error variance. Two specific solutions for analyzing the data are given.

In one solution it is assumed that the interaction is of the form $\gamma_{ij} = \lambda \alpha_i \gamma_j$ for every i and j. In this case maximum likelihood estimates of the parameters are obtained and a likelihood ratio test of Ho: $\lambda = 0$ is derived. The distribution of the likelihood ratio test is discussed and an estimate of the error variance is proposed when there is non-additivity in the data.

The second solution is a completely general procedure in which a theory of two by two table differences is presented and then used to propose estimates of the error variance that have desirable properties. Two examples illustrating the procedures are also presented.

Dallas Eugene Johnson
Department of Mathematics
and Statistics
Colorado State University
Fort Collins, Colorado 80521
December, 1970

ACKNOWLEDGEMENTS

A special thanks is given to Professor F. A. Graybill for his suggestion of the research areas, his patience and help with the research and with the dissertation, and his enthusiasm for statistics which is contagious. I am grateful to Professor D. C. Boes for serving on my committee and for his clear demonstration of excellence in teaching. I am also grateful to Professor P. W. Mielke, Professor R. E. Gaines, and the other members of the Department of Mathematics and Statistics for their time, effort, and encouragement throughout my graduate program.

I am deeply indebted to my wife, Erma Jean, and my son, Kelly, for their patience, love, and understanding during the last four years, as many, more hours were spent away from them than were spent with them.

Thanks are also due the National Institute of Health for its financial support during my graduate studies.

TABLE OF CONTENTS

Abstract of Dissertation iii

Acknowledgement iv

1. Introduction 1

1.1 Statement of the Problem 1
1.2 Literature Review 2

2. A General Method of Estimating σ^2 5

2.1 The Theory of Two by Two Tables 5
2.2 An Upper Bound on Two by Two Table Differences 15
2.3 Distribution of the Nonzero Roots of $\underline{Z}'\underline{Z}$ 29
2.4 Determining Significance of Two by Two Table Differences 35
2.5 Estimates of σ^2 and Hypothesis Testing 44
2.6 An Example 54

3. The Model $y_{ij} = \mu + \tau_i + \beta_j + \lambda \alpha_i \gamma_j + e_{ij}$ 59

3.1 Introduction 59
3.2 Maximum Likelihood Estimates 64
3.3 Likelihood Ratio Test of Ho: $\lambda = 0$ vs. Ha: $\lambda \neq 0$ 69

4. More On the Distribution of $\ell_1 / (\ell_1 + \ell_2 + \cdots + \ell_{b-1})$ 76

4.1 Introduction 76
4.2 The Special Case $b = 3$ 76
4.3 The Special Case $b = 4$ 87
4.4 An Approximation to the Distribution of $\ell_1 / (\ell_1 + \ell_2 + \cdots + \ell_{b-1})$ 102
4.5 An Estimate of σ^2 107
4.6 An Example 108

5. Problems Yet To Be Solved 110

Appendix 1 113

Bibliography 114

CHAPTER 1

INTRODUCTION

1.1 Statement of the Problem

In experimental situations a data analyst is sometimes faced with a two-way classification model without replication in which there is some evidence of non-additivity. Up to now the method of analysis has been to test for non-additivity by some available test such as Tukey's single degree of freedom test for non-additivity. If one then concludes that the data are additive, well known methods are available for completing the analysis of the data. However, if one concludes that the data are non-additive, the only recommendations have been to find a transformation which will make the transformed data additive.

It seems that there is merit in leaving the data in the units in which it was originally hoped to do the analysis, if only for ease of interpretation. In this instance the analyst would generally require an unbiased estimate of the error variance in order to make inferences about the main effects and would generally like to know what the pattern of interaction is. In view of the above remarks, two alternate procedures of analyzing the data are suggested.

The first is a completely general solution in that it may be used for any two-way classification model with one observation per cell and with any interaction structure. In this case a method for obtaining an estimate of σ^2 is given and conditional tests of hypotheses about main effects are proposed.

The second solution assumes a somewhat general model with a particular type of interaction structure. In this model a test for interaction is proposed and an estimate of the error variance is proposed when the data shows evidence of non-additivity.

1.2 Literature Review

There are several available tests for non-additivity in the two-way classification model with one observation per cell. The first test for non-additivity in this model was proposed by Tukey (1949). Tukey partitioned one degree of freedom for non-additivity from the residual sum of squares. Ward and Dick (1952) discussed an iterative solution to the set of normal equations one would obtain by considering the model

$$(1.2.1)\quad y_{ij} = \mu + \tau_i + \beta_j + \alpha\tau_i\beta_j + e_{ij} \quad i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, b.$$

They showed that the sum of squares due to the hypothesis that $\alpha = 0$ in (1.2.1) after one iteration was the same as the sum of squares for non-additivity proposed by Tukey. Mandel (1961) constructed a test for non-additivity for the two way classification model

$$(1.2.2)\quad y_{ij} = \mu + \tau_i + \beta_j + \alpha_i\beta_j + e_{ij} \quad i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, b.$$

Tukey (1962) added another stage to this development when he considered the model

$$(1.2.3)\quad y_{ij} = \mu + \tau_i + \beta_j + \alpha_i\beta_j + \tau_i\gamma_j + e_{ij} \quad i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, b$$

which was termed the "vacuum cleaner." Milliken (1969) and Milliken and Graybill (1970) proposed an extended general linear model to be used for testing for interaction and showed that the models (1.2.1) and (1.2.2) are special cases of their model. The model they considered was

(1.2.4) $\underline{y} = X\underline{\beta} + F\underline{\alpha} + e$ where F is a matrix whose elements are known functions of the elements in $X\underline{\beta}$.

Tests for interaction have been proposed for the latin square design by Mandel (1959), Tukey (1955), and Milliken (1969). Ward an Dick (1952) gave an iterative solution to the normal equations one would obtain from a Balanced Incomplete Block design when the interaction is assumed to be a product of the treatment and block effects. Milliken and Graybill (1970) also considered the Balanced Incomplete Block design. The model proposed by Milliken (1969) and Milliken and Graybill (1970) is a general model which can be applied to any design with any type of interaction structure as long as the interaction terms are known functions of the various treatment sets. Williams (1952) considered the model

$$(1.2.5) \qquad y_{ijk} = \mu + \beta_j + \alpha_i \gamma_j + e_{ijk} \quad i = 1, 2, \ldots, t;$$

$$j = 1, 2, \ldots, b; \; k = 1, 2, \ldots, n \text{ with } n > 1 .$$

The purpose of proposing this model was to aid in interpreting interaction effects. Gollob (1968) combines some features of factor analytic techniques and analysis of variance techniques when considering the two-way classification model. He calls his model the FANOVA model. The basic FANOVA model is essentially a two-way analysis of variance model which requires that the matrix of interaction parameters (γ_{ij}) be expressed as the sum of several successive multiplicative contrasts such that each contrast is orthogonal to all previous contrasts and accounts for a maximum of the remaining variance of the (γ_{ij}). Mandel (1969) developed the same basic model as that of Gollob. A basic difference in Gollob's method and Mandel's method is the number of degrees of freedom to be assigned to each of the partitions of the interaction sum of squares. The results of this paper supports Mandel's approach. Another basic difference in the two papers is that Gollob assumes a replicated

experiment and thus has an independent estimate of the error variance; whereas, Mandel assumes one observation per cell.

Several other references have been used and the extent of the contributions will be subsequently discussed.

CHAPTER 2

A GENERAL METHOD OF ESTIMATING σ^2

2.1 The Theory of Two by Two Tables.

The model for a two-way classification analysis of variance with interaction and without replication is

$$(2.1.1)\quad y_{ij} = \mu_{ij} + e_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + e_{ij} \quad i = 1, 2, \ldots, t;$$

$$j = 1, 2, \ldots, b$$

where y_{ij} is an observable random variable, μ is an unknown parameter defined to be the general mean, $\tau_1, \tau_2, \ldots, \tau_t$ is one set of unknown parameters defined to be the treatment effects, $\beta_1, \beta_2, \ldots, \beta_b$ is a second set of unknown parameters defined to be block effects, and $\gamma_{11}, \gamma_{12}, \ldots, \gamma_{tb}$ is a third set of unknown parameters defined to be the interaction or the non-additive effects of treatments by blocks. The effects defined to be block effects are nothing more than another set of treatment effects and called block effects only to simplify the discussion. The usual assumptions concerning the parameters will be made in order to insure estimability. These are

$$(2.1.2)\quad \sum_{i=1}^{t} \tau_i = \sum_{j=1}^{b} \beta_j = 0 \quad \text{and}$$

$$(2.1.3)\quad \sum_{i=1}^{t} \gamma_{ij} = 0 \text{ for every } j \text{ and } \sum_{j=1}^{b} \gamma_{ij} = 0 \text{ for every } i .$$

It is also assumed that the e_{ij} $i = 1, 2, \ldots, t$; $j = 1, 2, \ldots, b$ are unobservable random variables independently and identically distributed as normal random variables with zero mean and variance, σ^2 (i.e. $e_{ij} \sim$ i.i.d. $N(0, \sigma^2)$ $i = 1, 2, \ldots, t$; $j = 1, 2, \ldots, b$).

In the above model the uniformly minimum variance unbiased estimators of the parameters are

$$(2.1.4)\quad \hat{\mu} = y_{..}\,,\ \hat{\tau}_i = y_{i.} - y_{..} \quad i = 1, 2, \ldots, t,\ \hat{\beta}_j = y_{.j} - y_{..} \quad j = 1, 2, \ldots, b,$$

$$\hat{\gamma}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..} \quad i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, b\ .$$

The dot notation used above indicates that the subscript replaced by a dot has been summed over and the arithmetic mean obtained;

i.e. $y_{i.} = \frac{1}{b} \sum_{j=1}^{b} y_{ij}$. Before now there has been no known unbiased estimate of σ^2, unless all of the γ_{ij} are zero. In this case the estimate is

$$(2.1.5)\quad \hat{\sigma}^2 = \left[\sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2 \right] \Big/ (b-1)(t-1)$$

which is uniformly minimum variance unbiased. If the γ_{ij} are not all zero the expected value of the above estimate is

$$E\hat{\sigma}^2 = \left[\sum_{i=1}^{t}\sum_{j=1}^{b} E(y_{ij} - y_{i.} - y_{.j} + y_{..})^2\right] \Big/ (b-1)(t-1)$$

$$= \left[\sum_{i=1}^{t}\sum_{j=1}^{b} E(\gamma_{ij} + e_{ij} - e_{i.} - e_{.j} + e_{..})^2\right] \Big/ (b-1)(t-1)$$

$$= \left[\sum_{i=1}^{t}\sum_{j=1}^{b} (\gamma_{ij}^2 + E(e_{ij} - e_{i.} - e_{.j} + e_{..})^2)\right] \Big/ (b-1)(t-1)$$

$$= \left[\sum_{i=1}^{t}\sum_{j=1}^{b} \gamma_{ij}^2 + \sum_{i=1}^{t}\sum_{j=1}^{b} \frac{(b-1)(t-1)}{bt}\sigma^2\right] \Big/ (b-1)(t-1)$$

$$(2.1.6) \qquad = \sigma^2 + \frac{1}{(b-1)(t-1)} \sum_{i=1}^{t}\sum_{j=1}^{b} \gamma_{ij}^2 .$$

For further discussion and justification of the above statements see Graybill (1961).

The problem is to find a suitable estimate of σ^2 when the γ_{ij} are not all zero. First the idea of interaction will be reformulated by the use of two by two table differences. Let

$$(2.1.7) \qquad \mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} \quad i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, b .$$

A two by two table difference will mean

$$(2.1.8) \qquad \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'}$$

for some i, $i' = 1, 2, \ldots, t$ and j, $j' = 1, 2, \ldots, b$ with $i \neq i'$ and $j \neq j'$.

The following theorem gives necessary and sufficient conditions for all of the γ_{ij} to be equal to zero.

Theorem 2.1.1 Let $\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$ $i = 1, 2, \ldots, t$; $j = 1, 2, \ldots, b$ where the τ's, β's, and γ's satisfy the restrictions given in (2.1.2) and (2.1.3). A necessary and sufficient condition that $\gamma_{ij} = 0$ for every i and j is that $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$ for every i, i', j, and j'.

Proof:

$$\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = (\mu + \tau_i + \beta_j + \gamma_{ij}) - (\mu + \tau_{i'} + \beta_j + \gamma_{i'j})$$

$$(2.1.9) \qquad - (\mu + \tau_i + \beta_{j'} + \gamma_{ij'}) + (\mu + \tau_{i'} + \beta_{j'} + \gamma_{i'j'})$$

$$= \gamma_{ij} - \gamma_{i'j} - \gamma_{ij'} + \gamma_{i'j'}$$

Therefore, if $\gamma_{ij} = 0$ for all i and j it follows that $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$ for every i, i', j, and j'. Now $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$ for every i, i', j, and j' implies $\sum_{i'=1}^{t} \sum_{j'=1}^{b} (\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'}) = 0$ for every i and j. Thus, by (2.1.9), one obtains $\sum_{i'=1}^{t} \sum_{j'=1}^{b} (\gamma_{ij} - \gamma_{i'j} - \gamma_{ij'} + \gamma_{i'j'}) = 0$ which implies that $bt\,\gamma_{ij} = 0$ for every i and j and the result follows.

Definition 2.1.1 The two-way classification model

$$(2.1.1) \qquad y_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + e_{ij} = \mu_{ij} + e_{ij}$$

will be defined to be an additive model if and only if $\gamma_{ij} = 0$ for every

i and j, or equivalently, if $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$ for every i, i', j and j'. In the contrary case it will be defined to be a non-additive model, or a model with interaction.

Theorem 2.1.2 In the additive model the uniformly minimum variance unbiased estimate of σ^2 given by (2.1.5) is equal to

$$(2.1.10) \qquad \frac{\sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{\substack{i'=1 \\ i' \neq i}}^{t}\sum_{\substack{j'=1 \\ j' \neq j}}^{b}(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4bt(b-1)(t-1)}$$

Proof:

First note that if $i = i'$ or $j = j'$, then $y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'} = 0$. Thus (2.1.10) is equal to

$$(2.1.11) \qquad \frac{\sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{i'=1}^{t}\sum_{j'=1}^{b}(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4bt(b-1)(t-1)} .$$

Now the numerator of (2.1.11) is equal to

$$\sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{i'=1}^{t}\sum_{j'=1}^{b}(y_{ij}^2 + y_{i'j}^2 + y_{ij'}^2 + y_{i'j'}^2 - 2y_{ij}y_{i'j} - 2y_{ij}y_{ij'}$$

$$+ 2y_{ij}y_{i'j'} + 2y_{i'j}y_{ij'} - 2y_{i'j}y_{i'j'} - 2y_{ij'}y_{i'j'})$$

$$= bt\sum_{i=1}^{t}\sum_{j=1}^{b}y_{ij}^2 + bt\sum_{i'=1}^{t}\sum_{j=1}^{b}y_{i'j}^2 + bt\sum_{i=1}^{t}\sum_{j'=1}^{b}y_{ij'}^2$$

$$+ bt\sum_{i'=1}^{t}\sum_{j'=1}^{b}y_{i'j'}^2 - 2bt^2\sum_{j=1}^{b}y_{\cdot j}^2 - 2b^2t\sum_{i=1}^{t}y_{i\cdot}^2 + 2b^2t^2y_{\cdot\cdot}^2$$

$$+ 2b^2t^2y_{..}^2 - 2b^2t\sum_{i'=1}^{t} y_{i'.}^2 - 2bt^2\sum_{j'=1}^{b} y_{.j'}^2$$

$$= 4bt\left(\sum_{i=1}^{t}\sum_{j=1}^{b} y_{ij}^2 - b\sum_{i=1}^{b} y_{i.}^2 - t\sum_{j=1}^{b} y_{.j}^2 + bty_{..}^2\right)$$

$$= 4bt\left(\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2\right)$$

and the result follows.

In view of the above theorem observe that the uniformly minimum variance unbiased estimator of σ^2 can be written in terms of the two by two table differences.

The advantage of considering the problem in terms of two by two tables is that it is possible for some of the two by two table differences to be zero while none of the γ_{ij} are zero. This is illustrated in the following example.

Example 2.1.1 Consider the following 3 × 3 design in which the entries are the true cell means; that is, the μ_{ij}.

	B_1	B_2	B_3	Totals
T_1	0	2	4	6
T_2	2	4	6	12
T_3	7	0	11	18
Totals	9	6	21	36

Observe that $\mu = 4$, $\tau_1 = -2$, $\tau_2 = 0$, $\tau_3 = 2$, $\beta_1 = -1$, $\beta_2 = -2$, $\beta_3 = 3$, $\gamma_{11} = -1$, $\gamma_{12} = 2$, $\gamma_{13} = -1$, $\gamma_{21} = -1$, $\gamma_{22} = 2$, $\gamma_{23} = -1$, $\gamma_{31} = 2$, $\gamma_{32} = -4$, $\gamma_{33} = 2$ while $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$, $\mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} = 0$,

$\mu_{12} - \mu_{13} - \mu_{22} + \mu_{23} = 0$, and $\mu_{21} - \mu_{23} - \mu_{31} + \mu_{33} = 0$. Thus none of the γ_{ij} are zero while several of the two by two table differences are zero.

Definition 2.1.2 The statistic $\underline{y}' A \underline{y}/k$ will be defined to be a k degree of freedom estimate of σ^2 if and only if $\underline{y}' A \underline{y} \sim \sigma^2 \cdot \chi^2(k)$ where $\underline{y}' = [y_{11}y_{12}\cdots y_{1b}y_{21}y_{22}\cdots y_{2b}\cdots y_{t1}y_{t2}\cdots y_{tb}]$ is the vector of observations and A is a $bt \times bt$ known constant matrix.

Theorem 2.1.3 If a two by two table difference $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'}$ is zero, then $(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2/4$ is a one degree of freedom estimate of σ^2.

Proof:

First note that without any loss of generality one can assume that $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} = 0$. Now

$(y_{11} - y_{12} - y_{21} + y_{22})^2/4 = \underline{y}' \underline{a}\,\underline{a}' \underline{y}/4$ where

$\underline{a}' = [1 -1\ 0\cdots 0\ -1\ 1\ 0\cdots 0\ 0\cdots 0]$. By Corollary 4.7.1, Graybill (1961), one needs only to show that $\underline{a}\,\underline{a}'/4$ is idempotent of rank 1 and that the noncentrality parameter $\lambda = \frac{1}{2\sigma^2}\, \underline{\mu}' \underline{a}\,\underline{a}' \underline{\mu} = 0$ where

$\underline{\mu}' = E\underline{y}' = [\mu_{11} - \mu_{12}\cdots\mu_{1b}\mu_{21}\mu_{22}\cdots\mu_{2b}\cdots\mu_{t1}\mu_{t2}\cdots\mu_{tb}]$.

$\frac{\underline{a}\,\underline{a}'}{4} \cdot \frac{\underline{a}\,\underline{a}'}{4} = \frac{1}{4}\,\underline{a}\,\frac{(\underline{a}'\underline{a})}{4}\,\underline{a}' = \frac{\underline{a}\,\underline{a}'}{4}$ since $\underline{a}'\underline{a} = 4$. Hence $\frac{\underline{a}\,\underline{a}'}{4}$ is idempotent so that rank $\frac{(\underline{a}\,\underline{a}')}{4} = \mathrm{tr}\left(\frac{\underline{a}\,\underline{a}'}{4}\right) = \frac{1}{4}\,\mathrm{tr}\,(\underline{a}'\underline{a}) = 1$.

$\underline{a}'\,\underline{\mu} = 0$ by the hypothesis which implies $\lambda = 0$. Therefore

$(y_{11} - y_{12} - y_{21} + y_{22})^2/4 \sim \sigma^2 \chi^2(1)$ which implies that

$(y_{11} - y_{12} - y_{21} + y_{22})^2/4$ is a one degree of freedom estimate of σ^2.

In Example 2.2.1 note that the three possible two by two table differences involving treatments one and two are each zero. In this particular case one could consider only this portion of the design and get a two degree of freedom estimate of σ^2. This is generalized in the next theorem.

Theorem 2.1.4 If all of the two by two table differences involving the treatments $i_1, i_2, \ldots, i_k$, $2 \leq k \leq t$ and blocks $j_1, j_2, \ldots, j_m$, $2 \leq m \leq b$ are zero, then by considering only that portion of the design one can obtain a $(k-1)(m-1)$ degree of freedom estimate of σ^2. The estimate is

$$\frac{1}{(k-1)(m-1)} \sum_{p=1}^{k} \sum_{q=1}^{m} \left(y_{i_p j_q} - \frac{1}{k} \sum_{p=1}^{k} y_{i_p j_q} - \frac{1}{m} \sum_{q=1}^{m} y_{i_p j_q} + \frac{1}{mk} \sum_{p=1}^{k} \sum_{q=1}^{m} y_{i_p j_q} \right)^2 . \tag{2.1.12}$$

Proof:

Without any loss of generality it can be assumed that every two by two table difference in the portion of the design consisting of the first k treatments and the first m blocks is zero. Now (2.1.12) is equal to

$$\frac{1}{(k-1)(m-1)} \underline{y}' \begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix} \underline{y}$$

$$= \frac{1}{(k-1)(m-1)} \underline{y}' A \underline{y} \text{ (say)} \tag{2.1.13}$$

where $B \otimes C$ represents the direct product of the two matrices B and C.

$$A^2$$

$$= \left(\begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix} \right) \left(\begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \right) \otimes \left(\begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix} \right)$$

$$= \begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix} = A$$

Hence A is idempotent so that

$$\text{rank}(A) = \text{tr}(A)$$

$$= \text{tr} \begin{bmatrix} I_m - \frac{1}{m} J_m & 0 \\ 0 & 0 \end{bmatrix} \cdot \text{tr} \begin{bmatrix} I_k - \frac{1}{k} J_k & 0 \\ 0 & 0 \end{bmatrix}$$

$$= (m-1)(k-1) .$$

The noncentrality parameter is

$$\lambda = \frac{1}{2\sigma^2} \underline{\mu}' A \underline{\mu}$$

(2.1.14)

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{m} \left(\mu_{ij} - \frac{1}{k} \sum_{i'=1}^{k} \mu_{i'j} - \frac{1}{m} \sum_{j'=1}^{m} \mu_{ij'} + \frac{1}{mk} \sum_{i'=1}^{k} \sum_{j'=1}^{m} \mu_{i'j'} \right)^2$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{m} \left[\frac{1}{mk} \sum_{i'=1}^{k} \sum_{j'=1}^{m} \left(\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} \right) \right]^2$$

The quantity inside the brackets is zero since $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$

for all i, i'= 1, 2, . . . , k and j, j' = 1, 2, . . . , m by the hypothesis. Thus $\lambda = 0$ and by Corollary 4.7.1, Graybill (1961),

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\left(y_{ij} - \frac{1}{k}\sum_{i=1}^{k} y_{ij.} - \frac{1}{m}\sum_{j=1}^{m} y_{ij} + \frac{1}{mk}\sum_{i=1}^{k}\sum_{j=1}^{m} y_{ij}\right)^2$$

$$\sim \sigma^2 \cdot \chi^2\left((k-1)(m-1)\right)$$

and the result follows.

The following theorem is a generalization of the above theorem.

Theorem 2.1.5 Let $\xi_1, \xi_2, \ldots, \xi_n$ denote all of the distinct two by two table differences which are zero. Let A be the $n \times bt$ constant matrix such that $\xi' = [\xi_1, \xi_2, \cdots, \xi_n]' = \underline{\mu}' A'$ where $\underline{\mu}' = [\mu_{11}, \mu_{12}, \ldots, \mu_{tb}]$. If the rank of A is k, then $\underline{y}' A^- A \underline{y}$ is a k degree of freedom estimate of σ^2 where A^- indicates the Penrose generalized inverse of the matrix A as defined in Penrose (1955).

Proof: A^-A is idempotent by the definition of the generalized inverse. The rank $(A^-A) = k$ since $k = \text{rank}(A) \geq \text{rank}(A^- A) \geq \text{rank}(A A^- A) = \text{rank}(A) = k$. $\lambda = \frac{1}{2\sigma^2}\underline{\mu}' A^- A \underline{\mu} = 0$ since $A\underline{\mu} = \underline{0}$ by the hypothesis. Therefore by Corollary 4.7.1, Graybill (1961), $\underline{y}' A^- A \underline{y} \sim \sigma^2 \cdot \chi^2(k)$ and the result follows.

Given any two by two table there are two possible differences that one could write down. For example, consider the two by two table formed by the first two treatments and the first two blocks. The two possible differences that one could write are $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$ and $-\mu_{11} + \mu_{12} + \mu_{21} - \mu_{22}$. It will now be shown that the estimate of σ^2 given in Theorem 2.1.4 is invariant with respect to permutations of the two by two tables and also invariant with respect to either of the two possible differences which may be chosen for each two by two table.

Theorem 2.1.6 Let P be any $n \times n$ permutation matrix and D be any diagonal matrix with plus or minus ones on the diagonal. Then $(PDA)^{-}(PDA) = A^{-}A$.

Proof: Since P and D are orthogonal matrices, PD is an orthogonal matrix and hence by Theorem 6.2.10, Graybill (1969)

$$(PDA)^{-}(PDA) = A^{-}(PD)'(PD)A$$

$$= A^{-}D'P'PDA$$

$$= A^{-}A .$$

Corollary 2.1.1 Let $\underline{\xi} = A\underline{\mu}$ and $\underline{\xi}^{*} = PDA\underline{\mu}$ where P and D are the same as in Theorem 2.1.6, then $\underline{y}'A^{-}A\underline{y} = \underline{y}'(PDA)^{-}(PDA)\underline{y}$.

Proof: Note that $A^{-}A = (PDA)^{-}(PDA)$ from Theorem 2.1.6 and the result follows.

The above corollary is important since it guarantees that no matter which order one writes down the contrasts or which of the two possible differences one uses, the estimate of σ^2 obtained is unique. That is, every data analyst using the techniques that follow on the same set of data will necessarily obtain the same estimate of σ^2.

In the following sections of this chapter a procedure will be given to determine which if any of the estimates of two by two table differences are significantly different from zero. Then σ^2 will be estimated from those differences not significantly different from zero.

2.2 An Upper Bound on Two by Two Table Differences

In view of the last section, it is seen that if one were able to determine which, if any, of the two by two table differences $\xi(i, i', j, j') = \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'}$ are zero, an unbiased estimate of σ^2 could be found. In order to determine if σ^2 can be estimated in this way, a procedure will be given to simultaneously test the hypothesis that all two by two table differences are zero. To do this one would like to determine the distribution of

$Q = \max_{i,i',j,j'} \frac{(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4}$ under the hypothesis that all two by two table differences are zero the distribution of this statistic is not known and it appears that finding this distribution is a very difficult problem. Therefore it seems that the next best thing to do is to find the distribution of an upper bound to the statistic Q. There are, of course, many upper bounds to the statistic Q. Ideally one would like an upper bound that is quite close to Q. Also one would like an upper bound whose distribution does not depend on any of the parameters other than σ^2 if all two by two table differences are zero. Note that if all two by two table differences are zero, the distribution of $\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2$ depends on no parameters other than σ^2 since $\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2 \sim \sigma^2 \cdot \chi^2((b-1)(t-1))$. It will be shown that $Q \leq \sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2$. Therefore it seems that a test function could be obtained by partitioning the residual sum of squares $\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2$ into two parts R_1 and R_2 satisfying $R_1 + R_2 = \sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2$ where $Q \leq R_1$ with R_1 close to Q.

Definition 2.2.1 A contrast of the observations is defined to belong to the error space if the expected value of that contrast is zero for all values of the parameters. Such a contrast will be defined to be normalized if the variance of the contrast is σ^2.

The following notation will be used:

$$\xi(i, i', j, j') = \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'}$$

and

$$\hat{\xi}(i, i', j, j') = y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'} .$$

Theorem 2.2.1 If $\xi(i, i', j, j') = 0$ for every i, i', j, j' in the model (2.1.1) then $E\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right) = 0$ if and only if $\sum_{i=1}^{t} c_{ij} = 0$ for every j and $\sum_{j=1}^{b} c_{ij} = 0$ for every i. For such contrasts $\operatorname{var}\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right) = \sigma^2$ if and only if $\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij}^2 = 1$.

Proof: By Theorem 2.1.1 $\xi(i, i', j, j') = 0$ for every i, i', j, j' if and only if $\gamma_{ij} = 0$ for every i and j. Thus

$$E\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right) = \sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} (\mu + \tau_i + \beta_j)$$

$$= \mu \sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} + \sum_{i=1}^{t} \tau_i \sum_{j=1}^{b} c_{ij}$$

$$+ \sum_{j=1}^{b} \beta_j \sum_{i=1}^{t} c_{ij}$$

so that $E\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right) = 0$ if and only if $\sum_{j=1}^{b} c_{ij} = 0$ for every i and $\sum_{i=1}^{t} c_{ij} = 0$ for every j.

$$\text{var}\left(\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}y_{ij}\right) = \text{var}\left[\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}(\mu + \tau_i + \beta_j + e_{ij})\right]$$

$$= \text{var}\left(\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}e_{ij}\right)$$

$$= \sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}^2\,\sigma^2 \quad \text{since } e_{ij} \sim \text{i.i.d.N}(0, \sigma^2).$$

Hence $\text{var}\left(\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}y_{ij}\right) = \sigma^2$ if and only if $\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}^2 = 1$.

A series of theorems and corollarys will now be stated and proved. These will be needed in the proofs of Theorem 2.2.4 and Theorem 2.2.5 which are concerned with the maximum value of $\left(\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ij}y_{ij}\right)^2$ when restrictions are placed on the c_{ij}.

Theorem 2.2.2 Let A be any $n \times n$ symmetric matrix with characteristic roots $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and corresponding characteristic vectors $\underline{p}_1, \underline{p}_2, \cdots, \underline{p}_n$ satisfying $\underline{p}_i'\underline{p}_i = 1$ for $i = 1, 2, \cdots, n$ and $\underline{p}_i'\underline{p}_j = 0$ for $i \neq j = 1, 2, \cdots, n$. The maximum value of $\underline{x}'A\underline{x}$ over all vectors $\underline{x}$ satisfying $\underline{x}'\underline{x} = 1$ is λ_1 and is attained when $\underline{x} = \underline{p}_1$.

Proof: Let $P = [\underline{p}_1\,\underline{p}_2 \cdots \underline{p}_n]$.

Then $P'AP = \begin{bmatrix} \underline{p}_1' \\ \underline{p}_2' \\ \vdots \\ \underline{p}_n' \end{bmatrix} A\,[\underline{p}_1\,\underline{p}_2 \cdots \underline{p}_n]$

$$= \text{Diag}(\lambda_1, \lambda_2, \cdots \lambda_n) = D(\text{say}).$$

For a given vector $\underline{x}$ satisfying $\underline{x}'\underline{x} = 1$, let $\underline{u} = P'\underline{x}$ which is equivalent to $\underline{x} = P\underline{u}$. Also $\underline{x}'\underline{x} = 1$ implies $\underline{u}'P'P\underline{u} = 1$ which implies $\underline{u}'\underline{u} = 1$. Therefore

$$\max_{\underline{x},\, \underline{x}'\underline{x}=1} = \max_{\underline{u},\, \underline{u}'\underline{u}=1} \underline{u}'P'AP\underline{u}$$

$$= \max_{\underline{u},\, \underline{u}'\underline{u}=1} \underline{u}'D\underline{u}$$

$$= \max_{\underline{u},\, \underline{u}'\underline{u}=1} (u_1^2\lambda_1 + u_2^2\lambda_2 + \cdots u_n^2\lambda_n)$$

Since $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, $u_1^2\lambda_1 + u_2^2\lambda_2 + \cdots + u_n^2\lambda_n$ will be a maximum when $u_1^2 = 1$ and $u_2^2 = u_3^2 = \cdots = u_n^2 = 0$. Hence $\max_{\underline{x},\, \underline{x}'\underline{x}=1} \underline{x}'A\underline{x} = \lambda_1$. Now $\underline{u}' = [1\,0\,0\cdots 0]$ implies $\underline{x} = P\underline{u} = \underline{p}_1$ so that the maximum is attained when $\underline{x} = \underline{p}_1$.

Theorem 2.2.3 Let A be any $n \times n$ symmetric matrix. Let B be any $n \times k$ matrix with $n \geq k$. The maximum value of $\underline{x}'A\underline{x}$ subject to the restrictions $B'\underline{x} = 0$ is equal to the largest characteristic root of $(I - BB^-)A(I - BB^-)$ and is attained when $\underline{x}$ is proportional to the characteristic vector of $(I - BB^-)A(I - BB^-)$ corresponding to the largest characteristic root of $(I - BB^-)A(I - BB^-)$.

Proof: Once it is shown that

$$(2.2.1) \qquad \max_{\substack{\underline{x} \neq \underline{0} \\ \underline{x}'B = \underline{0}'}} \frac{\underline{x}'A\underline{x}}{\underline{x}'\underline{x}} = \max_{\underline{x} \neq \underline{0}} \frac{\underline{x}'(I - BB^-)A(I - BB^-)\underline{x}}{\underline{x}'\underline{x}},$$

the result follows from Theorem 2.2.2. Now

$$\max_{\substack{\underline{x}'B=\underline{0}' \\ \underline{x}\neq\underline{0}}} \frac{\underline{x}'A\underline{x}}{\underline{x}'\underline{x}} = \max_{\underline{x}\neq\underline{0}} \frac{[\underline{x}'(I-BB^-)]\,A\,[(I-BB^-)\,\underline{x}]}{[\underline{x}'(I-BB^-)][(I-BB^-)\,\underline{x}]}$$

$$\geq \max_{\underline{x}\neq\underline{0}} \frac{\underline{x}'(I-BB^-)\,A\,(I-BB^-)\underline{x}}{\underline{x}'\underline{x}}$$

since BB^- non-negative implies $\underline{x}'\,BB^-\,\underline{x} \geq 0$ which implies $\underline{x}'\underline{x}' \geq \underline{x}'\underline{x}' - \underline{x}'\,BB^-\,\underline{x} = \underline{x}'(I-BB^-)\underline{x}$. Also

$$\max_{\substack{\underline{x}'B=\underline{0}' \\ \underline{x}\neq\underline{0}}} \frac{\underline{x}'A\underline{x}}{\underline{x}'\underline{x}} = \frac{\underline{x}^{*\prime}A\underline{x}^*}{\underline{x}^{*\prime}\underline{x}^*} \quad \text{where } \underline{x}^*$$

is a vector $\underline{x}$ that gives the expression on the left its maximum value. Now $\underline{x}^{*\prime}\underline{B}=\underline{0}'$ if and only if $\underline{x}^{*\prime} = \underline{x}^{*\prime}(I-BB^-)$. Hence

$$\frac{\underline{x}^{*\prime}A\underline{x}^*}{\underline{x}^{*\prime}\underline{x}^*} = \frac{\underline{x}^{*\prime}(I-BB^-)\,A\,(I-BB^-)\underline{x}^*}{\underline{x}^{*\prime}\underline{x}^*} \quad \text{and}$$

$$\frac{\underline{x}^{*\prime}(I-BB^-)\,A\,(I-BB^-)\underline{x}^*}{\underline{x}^{*\prime}\underline{x}^*} \leq \max_{\underline{x}\neq\underline{0}} \frac{\underline{x}'(I-BB^-)\,A\,(I-BB^-)\underline{x}}{\underline{x}'\underline{x}}\,.$$

Therefore the result follows.

Corollary 2.2.1 Let A and B be defined as in Theorem 2.2.3. The maximum value of $\underline{x}'A\underline{x}$ subject to the restrictions $B'\underline{x}=\underline{0}$ and $\underline{x}'\underline{x}=1$ is equal to the largest characteristic root of $(I-BB^-)A(I-BB^-)$ and is attained when $\underline{x} = \underline{p}_1$ where $\underline{p}_1$ is a characteristic vector of $(I-BB^-)A(I-BB^-)$ corresponding to the largest characteristic root of $(I-BB^-)A(I-BB^-)$ and satisfying $\underline{p}_1'\underline{p}_1 = 1$.

Proof:

$$\frac{\underline{x}'A\underline{x}}{\underline{x}'\underline{x}} = \frac{[(\underline{x}'\underline{x})^{-\frac{1}{2}}\underline{x}']\,A\,[(\underline{x}'\underline{x})^{-\frac{1}{2}}\underline{x}]}{[(\underline{x}'\underline{x})^{-\frac{1}{2}}\underline{x}'][(\underline{x}'\underline{x})^{-\frac{1}{2}}\underline{x}]}\,.$$

Thus for every vector $\underline{x}$ there exists a vector $\underline{u} = (\underline{x}'\underline{x})^{-\frac{1}{2}}\underline{x}$ satisfying $\underline{u}'\underline{u} = 1$ and the result follows from Theorem 2.2.3.

Corollary 2.2.2 Let A be any $n \times n$ symmetric matrix. The maximum value of $\underline{x}' A \underline{x}$ subject to the restrictions $\sum_{i=1}^{n} x_i = \underline{x}' \underline{j}_n = 0$ and $\sum_{i=1}^{n} x_i^2 = \underline{x}' \underline{x} = 1$ is the largest characteristic root of $(I - \frac{1}{n} J) A (I - \frac{1}{n} J)$ and is attained when $\underline{x} = \underline{p}_1$ where $\underline{p}_1$ is a characteristic vector of $(I - \frac{1}{n} J) A (I - \frac{1}{n} J)$ corresponding to the largest characteristic root of $(I - \frac{1}{n} J) A (I - \frac{1}{n} J)$ and satisfying $\underline{p}_1' \underline{p}_1 = 1$.

Proof: Note that $\underline{j}^- = \frac{1}{n} \underline{j}'$ and the results follows from Corollary 2.2.1.

Corollary 2.2.3 Let A be any symmetric $bt \times bt$ matrix. Let $\underline{c}' = [c_{11} c_{12} \cdots c_{1b} c_{21} c_{22} \cdots c_{2b} \cdots c_{t1} c_{t2} \cdots c_{tb}]$. The maximum value of $\underline{c}' A \underline{c}$ subject to the restrictions $\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij}^2 = 1$, $\sum_{i=1}^{t} c_{ij} = 0$ for every $j = 1, 2, \cdots, b$, and $\sum_{j=1}^{b} c_{ij} = 0$ for every $i = 1, 2, \cdots, t$ is λ_1, the largest characteristic root of

$$W = [(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)] A [(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)]$$

and is attained when $\underline{x} = \underline{p}_1$ where $\underline{p}_1$ is a characteristic vector of W corresponding to λ_1 and satisfying $\underline{p}_1' \underline{p}_1 = 1$.

Proof: $\sum_{i=1}^{t} c_{ij} = 0$ for every $j = 1, 2, \cdots, b$ is equivalent to $\underline{c}' (J_b \otimes I_t) = \underline{0}'$ and $\sum_{j=1}^{b} c_{ij} = 0$ for every $i = 1, 2, \cdots, t$ is equivalent to $\underline{c}' (I_b \otimes J_t) = \underline{0}'$. Also $\sum_{i=1}^{t} c_{ij} = 0$ for every $j = 1, 2, \cdots, b$ implies $\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} = 0$ which is equivalent to $\underline{c}' (J_b \otimes J_t) = \underline{0}'$. Now

(2.2.2) $\underline{c}'(J_b \otimes I_t) = \underline{0}'$, $\underline{c}'(I_b \otimes J_t) = \underline{0}'$, and $\underline{c}'(J_b \otimes J_t) = \underline{0}'$

if and only if

(2.2.3) $$\underline{c}'(\frac{1}{b} J_b \otimes I_t + I_b \otimes \frac{1}{t} J_t - \frac{1}{bt} J_b \otimes J_t) = \underline{0}' .$$

To prove this first note that

$$\frac{1}{b} J_b \otimes I_t + I_b \otimes \frac{1}{t} J_t - \frac{1}{bt} J_b \otimes J_t$$

is an idempotent matrix. Thus

$$\underline{c}'(\frac{1}{b} J_b \otimes I_t + I_b \otimes \frac{1}{t} J_t - \frac{1}{bt} J_b \otimes J_t) = \underline{0}'$$

if and only if

$$\underline{c}'(\frac{1}{b} J_b \otimes I_t + I_b \otimes \frac{1}{t} J_t - \frac{1}{bt} J_b \otimes J_t)\underline{c} = 0$$

which is equivalent to

(2.2.4) $$\underline{c}'(\frac{1}{b} J_b \otimes (I_t - \frac{1}{t} J_t) + I_b \otimes \frac{1}{t} J_t)\ \underline{c} = 0 .$$

Now both of the matrices $\frac{1}{b} J_b \otimes (I_t - \frac{1}{t} J_t)$ and $I_b \otimes \frac{1}{t} J_t$ are non-negative, hence $\underline{c}'(I_b \otimes \frac{1}{t} J_t)\underline{c} = 0$ which implies $\underline{c}'(I_b \otimes \frac{1}{t} J_t) = \underline{0}'$ since $I_b \otimes \frac{1}{t} J_t$ is idempotent. Therefore $\underline{c}'(I_b \otimes J_t) = \underline{0}'$. Similarly one can show that $\underline{c}'(J_b \otimes I_t) = \underline{0}'$. That the "only if" is true can be seen by direct substitution. Now to get the desired result one need only note that

$$B = (\frac{1}{b} J_b \otimes I_t + I_b \otimes \frac{1}{t} J_t - \frac{1}{bt} J_b \otimes J_t)$$

is idempotent and hence $B^- = B$ and $B B^- = B$. Thus

$$I - BB^{-} = I_{bt} - \frac{1}{b} J_b \otimes I_t - I_b \otimes \frac{1}{t} J_t + \frac{1}{bt} J_b \otimes J_t$$

$$= (I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)$$

and the result follows upon application of Corollary 2.2.1 .

Theorem 2.2.4 In the model (2.1.1) the maximum value of $\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right)^2$ subject to the restrictions $\sum_{i=1}^{t} c_{ij} = 0$ for $j = 1, 2, \cdots, b$, $\sum_{j=1}^{b} c_{ij} = 0$ for $i = 1, 2, \cdots, t$, and $\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij}^2 = 1$ is $\sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2$ and is attained when

$$c_{ij} = \frac{(y_{ij} - y_{i.} - y_{.j} + y_{..})}{\left[\sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - y_{i.} - y_{.j} + y_{.})^2 \right]^{\frac{1}{2}}} .$$

Proof: $\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right)^2 = (\underline{c}' \underline{y})^2 = \underline{c}' \underline{y} \underline{y}' \underline{c}$.

Therefore by Corollary 2.2.3 the maximum value of $\left(\sum_{i=1}^{t} \sum_{j=1}^{b} c_{ij} y_{ij} \right)^2$ subject to the restrictions given is

$$Ch_{max} \{ [(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)] \underline{y} \underline{y}' [(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)] \}$$

$$(2.2.5) \quad = Ch_{max} \{ \underline{y}' [(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)] \underline{y} \}$$

$$= \underline{y}' [(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t)] \underline{y}$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2 .$$

It can be seen that this value is attained for the c_{ij} given by direct substitution.

Note that the two by two table differences $(\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'})/2$ satisfy the above restrictions with $c_{ij} = \frac{1}{2}$, $c_{i'j} = -\frac{1}{2}$, $c_{ij'} = -\frac{1}{2}$, $c_{i'j'} = \frac{1}{2}$ and all other c's equal to zero. Thus

$$Q = \max_{i,i',j,j'} \frac{(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4} \leq \sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2 .$$

In order to get closer to the value of Q, additional restrictions will be placed on the c_{ij}. The new restrictions are $c_{ij} = a_i d_j$ with $\sum_{i=1}^{t} a_i = 0$, $\sum_{j=1}^{b} d_j = 0$, $\sum_{i=1}^{t} a_i^2 = 1$, and $\sum_{j=1}^{b} d_j^2 = 1$.

Theorem 2.2.5 The maximum value of $\left(\sum_{i=1}^{t} \sum_{j=1}^{b} a_i d_j y_{ij} \right)^2$ subject to the restrictions

$$(2.2.6) \qquad \sum_{i=1}^{t} a_i = \sum_{j=1}^{b} d_j = 0 \quad \text{and} \quad \sum_{i=1}^{t} a_i^2 = \sum_{j=1}^{b} d_j^2 = 1$$

is the largest characteristic root of $Z'Z$ where $Z = (I_t - \frac{1}{t} J_t) Y (I_b - \frac{1}{b} J_b)$ and Y is the $t \times b$ matrix of observations (i.e. $Y = (y_{ij})$). The maximum value is attained when $\underline{a}' = [a_1 a_2 \cdots a_t]$ is a characteristic vector of ZZ' corresponding to the largest characteristic root of ZZ' and satisfying $\underline{a}'\underline{a} = 1$ and $\underline{d}' = [d_1 d_2 \cdots d_b]$ is a characteristic vector of $Z'Z$ corresponding to the largest root of $Z'Z$ and satisfying $\underline{d}'\underline{d} = 1$.

Proof:

$$(2.2.7)\quad \sum_{i=1}^{t}\sum_{j=1}^{b} a_i d_j y_{ij} = (\underline{a}' Y \underline{d})^2 = \underline{a}' Y \underline{d}\underline{d}' Y' \underline{a} = \underline{d}' Y' \underline{a}\underline{a}' Y \underline{d} .$$

For each vector $\underline{d}$ satisfying (2.2.6), the maximum value of $\underline{a}' Y \underline{d}\underline{d}' Y' \underline{a}$ over all vectors $\underline{a}$ satisfying (2.2.6) is

$$(2.2.8)\quad \begin{aligned} &Ch_{max}\left[(I_t - \tfrac{1}{t} J_t) Y \underline{d}\underline{d}' Y' (I_t - \tfrac{1}{t} J_t)\right] \\ &= \underline{d}' Y' (I_t - \tfrac{1}{t} J_t) Y \underline{d} = \lambda_{\underline{d}} \text{ (say)} \end{aligned}$$

and is attained when $\underline{a}$ is a normalized characteristic vector of $(I_t - \frac{1}{t} J_t) Y \underline{d}\underline{d}' Y' (I_t - \frac{1}{t} J_t)$ corresponding to the root $\lambda_{\underline{d}}$, by Corollary 2.2.2. The maximum value of (2.2.8) over all vectors $\underline{d}$ satisfying (2.2.6) is

$$(2.2.9)\quad \begin{aligned} &Ch_{max}\left[(I_b - \tfrac{1}{b} J_b) Y' (I_t - \tfrac{1}{t} J_t) Y (I_b - \tfrac{1}{b} J_b)\right] \\ &= Ch_{max}(Z' Z) = \lambda \text{ (say)} \end{aligned}$$

and is attained when $\underline{d} = \underline{d}^*$ where $\underline{d}^*$ is a normalized characteristic vector of $Z' Z$ corresponding to the root λ. Hence

$$(2.2.10)\quad Z' Z \underline{d}^* = \lambda \underline{d}^*$$

and

$$(2.2.11)\quad \underline{d}^{*\prime} Z' Z \underline{d}^* = \lambda .$$

From above, the maximum value of $\underline{a}' Y \underline{d}^* \underline{d}^{*\prime} Y' \underline{a}$ is

$$(2.2.12)\quad \underline{d}^{*\prime} Y' (I_t - \tfrac{1}{t} J_t) Y \underline{d}^* = \lambda_{\underline{d}^*}$$

and is attained when $\underline{a} = \underline{a}^*$ where $\underline{a}^*$ is a normalized characteristic vector of

$$(2.2.13) \qquad (I_t - \frac{1}{t} J_t) Y \underline{d}^* \underline{d}^{*\prime} Y' (I_t - \frac{1}{t} J_t)$$

corresponding to the root $\lambda_{\underline{d}^*}$. Since $\underline{d}^{*\prime} \underline{j}_b = 0$, $\underline{d}^* = (I_b - \frac{1}{b} J_b)\underline{d}^*$. Substituting $(I_b - \frac{1}{b} J_b)\underline{d}^*$ for $\underline{d}^*$ in (2.2.12) one obtains

$$(2.2.14) \qquad \lambda_{\underline{d}^*} = \underline{d}^{*\prime} Z' Z \underline{d}^* = \lambda$$

from (2.2.11). Now $\underline{a}^*$ is a characteristic vector of (2.2.13), $\underline{d}^* = (I_b - \frac{1}{b} J_b)\underline{d}^*$, and (2.2.14) imply

$$(2.2.15) \qquad Z \underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^* = \lambda \underline{a}^*$$

which implies that

$$(2.2.16) \qquad Z' Z \underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^* = \lambda Z' \underline{a}^* .$$

Substituting $\lambda \underline{d}^*$ for $Z' Z \underline{d}^*$ in (2.2.16) one obtains $\lambda \underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^* = \lambda Z' \underline{a}^*$ which implies

$$(2.2.17) \qquad \underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^* = Z' \underline{a}^* .$$

Upon substitution of $Z' \underline{a}^*$ for $\underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^*$ in (2.2.15), one obtains

$$(2.2.18) \qquad Z Z' \underline{a}^* = \lambda \underline{a}^*$$

and hence $\underline{a}^*$ is a normalized characteristic vector of $Z Z'$ corresponding to the root λ.

It is interesting to note the following relationships among λ, $\underline{a}^*$, and $\underline{d}^*$ hold.

(2. 2. 19) $$\underline{a}^{*\prime} Z \underline{d}^* = \lambda^{\frac{1}{2}}$$

(2. 2. 20) $$\lambda^{\frac{1}{2}} \underline{d}^* = Z' \underline{a}^*$$

(2. 2. 21) $$\lambda^{\frac{1}{2}} \underline{a}^* = Z \underline{d}^* \ .$$

To prove (2. 2. 19) multiply each side of (2. 2. 15) on the left by $\underline{a}^{*\prime}$ to get $\underline{a}^{*\prime} Z \underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^* = \lambda$ which implies $\lambda = (\underline{a}' Z \underline{d}^*)^2$ and (2. 2. 19) follows. To prove (2. 2. 20), note that from (2. 2. 19) one has $\underline{d}^{*\prime} Z' \underline{a}^* = \lambda^{\frac{1}{2}}$. Multiplying each side of this on the left by $\underline{d}^*$ one obtains

(2. 2. 22) $$\lambda^{\frac{1}{2}} \underline{d}^* = \underline{d}^* \underline{d}^{*\prime} Z' \underline{a}^* = Z' \underline{a}^*$$

by (2. 2. 17). Multiplying each side of (2. 2. 20) on the left by Z one gets $\lambda^{\frac{1}{2}} Z d^* = Z Z' \underline{a}^* = \lambda \underline{a}^*$ by (2. 2. 18) and hence (2. 2. 21) follows. These relationships between the characteristic vectors of $Z Z'$ and $Z' Z$ are useful since one need only consider the matrix $Z Z'$ or the matrix $Z' Z$ which ever has the smaller dimension.

Note that the two by two table differences satisfy the restrictions (2. 2. 6) with $a_i = \frac{1}{\sqrt{2}}$, $a_i' = \frac{-1}{\sqrt{2}}$, $d_j = \frac{1}{\sqrt{2}}$, $d_j' = \frac{-1}{\sqrt{2}}$ and all other a's and d's equal to zero. Hence

$$Q = \max_{i,i',j,j'} \frac{(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4} \leq Ch_{max}(Z' Z) \ .$$

It is also interesting to note that Tukey's sum of squares for non-additivity satisfy the restrictions (2.2.6). Tukey's sum of squares for non-additivity is

$$(2.2.23) \qquad \frac{\left(\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{i.}-y_{..})(y_{.j}-y_{..})\,y_{ij}\right)^2}{\sum_{i=1}^{t}(y_{i.}-y_{..})^2\sum_{j=1}^{b}(y_{.j}-y_{..})^2}$$

$$= \left\{\sum_{i=1}^{t}\frac{y_{i.}-y_{..}}{\sqrt{\sum_{i=1}^{t}(y_{i.}-y_{..})^2}}\,\frac{y_{.j}-y_{..}}{\sqrt{\sum_{j=1}^{b}(y_{.j}-y_{..})^2}}\,y_{ij}\right\}^2.$$

Thus letting $a_i = \dfrac{y_{i.}-y_{..}}{\sqrt{\sum_{i=1}^{t}(y_{i.}-y_{..})^2}}$ $\quad i=1,2,\cdots,t$ and

$b_j = \dfrac{y_{.j}-y_{..}}{\sqrt{\sum_{j=1}^{b}(y_{.j}-y_{..})^2}}$ and therefore (2.2.23) is at most equal to the largest characteristic root of $Z'Z$.

To summarize let $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_b$ be the characteristic roots of $Z'Z$. Then

$$(2.2.24) \qquad Q \leq \ell_1 \leq \ell_1 + \ell_2 + \cdots + \ell_b = \mathrm{tr}(Z'Z) = \sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij}-y_{i.}-y_{.j}+y_{..})^2 .$$

It is interesting to note that there is no ordering between Q and Tukey's sum of squares for non-additivity.

In the next section the distribution of the nonzero characteristic roots of $Z'Z$ will be given.

2.3 Distribution of the Nonzero Roots of $Z'Z$

Throughout the rest of this study it will be assumed that $b \leq t$. If $b > t$ one needs only to interchange b and t in all of the results that follow.

A series of theorems and corollaries are now given leading to Corollary 2.3.4, which shows that the distribution of the non-zero characteristic roots of $Z'Z$ is the same as the distribution of the characteristic roots of a matrix having the Wishart distribution.

Theorem 2.3.1 Let A be any $n \times n$ symmetric idempotent matrix of rank k. Let H be any $n \times k$ matrix of rank k such that $A = HH'$. Then $H'H = I_k$ and $H^- = H'$.

Proof: $A^2 = A$ implies

$$HH'HH' = HH'$$

which implies

$$H^-(HH'HH')H'^- = H^-(HH')H'^-$$

which implies

$$(H^-H)(H'H)(H'H'^-) = (H^-H)(H'H'^-) \quad . \tag{2.3.1}$$

Now $H^-H = I_k$ and $H'H'^- = (H^-H)' = I_k' = I_k$

so that (2.3.1) becomes

$$I_k H'H I_k = I_k I_k$$

and hence

$$H'H = I_k \quad . \tag{2.3.2}$$

Multiplying both sides of (2.3.2) on the right by H^-, one obtains $H'HH^- = H^-$ which implies $H'H'^-H' = H^-$ and hence

(2.3.3) $$H' = H^{-}.$$

Corollary 2.3.1 Let K be any $n \times n-1$ matrix of rank $n-1$ such that $KK' = I_n - \frac{1}{n} J_n$. Then

(2.3.4) $$K'K = I_{n-1} \quad \text{and} \quad K^{-} = K'.$$

Proof: $I_n - \frac{1}{n} J_n$ is symmetric idempotent of rank $n-1$ and the results follow by Theorem 2.3.1.

Theorem 2.3.2 Let A and H be defined as in Theorem 2.3.1. Let B be any $n \times n$ (real) matrix. The nonzero characteristic roots of ABA are the same as the nonzero characteristic roots of $H'BH$.

Proof: Suppose λ is a nonzero characteristic root of ABA and that $\underline{x}$ is its corresponding characteristic vector. Then $ABA\underline{x} = \lambda \underline{x}$ which implies

(2.3.5) $$HH'BHH'\underline{x} = \lambda \underline{x}.$$

Multiplying both sides of (2.3.5) on the left by H' and applying (2.3.2) one obtains $H'BH(H'\underline{x}) = \lambda(H'\underline{x})$. Hence λ is a nonzero characteristic root of $H'BH$ corresponding to the characteristic vector $H'\underline{x}$. Now suppose that α is a nonzero characteristic root of $H'BH$ corresponding to the characteristic vector $\underline{u}$. Then

$$H'BH\underline{u} = \alpha \underline{u}$$

which implies

$$HH'BH\underline{u} = \alpha H\underline{u}$$

which implies

$$HH'BHH'H\underline{u} = \alpha H\underline{u}$$

since $H'H = I_k$. Thus $ABA(H\underline{u}) = \alpha(H\underline{u})$ and hence α is a nonzero

characteristic root of ABA corresponding to the characteristic vector $H\underline{u}$.

Corollary 2.3.2 Let K be any $b \times b-1$ matrix of rank $b-1$ such that $KK' = I_b - \frac{1}{b}J_b$. Let Y be the $t \times b$ matrix of observations in a $t \times b$ two-way classification model, then the nonzero roots of $(I_b - \frac{1}{b}J_b)\, Y'(I_t - \frac{1}{t}J_t)\, Y\, (I_b - \frac{1}{b}J_b)$ are the same as the nonzero characteristic roots of $K'Y'(I_t - \frac{1}{t}J_t)\, Y\, K$.

Proof: Let $B = Y'(I_t - \frac{1}{t}J_t)\, Y$ and $H = K$ in Theorem 2.3.2.

Theorem 2.3.3 If rank $(Y) = b$ with probability one, then rank $(Z) = b-1$ with probability one where $Z = (I_t - \frac{1}{t}J_t)\, Y\, (I_b - \frac{1}{b}J_b)$ and $t \geq b$.

Proof: It is clear that the rank $(Z) \leq b-1$ since $Z\underline{j}_b = \underline{0}$. Let $Y = [\underline{y}_1 \underline{y}_2 \cdots \underline{y}_b]$ and $Z = [\underline{z}_1 \underline{z}_2 \cdots \underline{z}_b]$, then $\underline{z}_j = \underline{y}_j - \underline{y}_.$ where $\underline{y}_. = \frac{1}{b}\sum_{j=1}^{b} \underline{y}_j$. Now suppose rank $(Z) < b-1$. This implies that there exist constants $a_1, a_2, \cdots, a_{b-1}$ not all zero such that $a_1\underline{z}_1 + a_2\underline{z}_2 + \cdots + a_{b-1}\underline{z}_{b-1} = \underline{0}$ since $\underline{z}_b = -(\underline{z}_1 + \underline{z}_2 + \cdots + \underline{z}_{b-1})$. Thus $a_1(\underline{y}_1 - \underline{y}_.) + a_2(\underline{y}_2 - \underline{y}_.) + \cdots + a_{b-1}(\underline{y}_{b-1} - \underline{y}_.) = \underline{0}$ which implies

$$\left(a_1 - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b}\right)\underline{y}_1 + \left(a_2 - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b}\right)\underline{y}_2$$

$$+ \cdots + \left(a_{b-1} - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b}\right)\underline{y}_{b-1} \tag{2.3.6}$$

$$+ \left(-\frac{a_1 + a_2 + \cdots + a_{b-1}}{b}\right)\underline{y}_b = \underline{0}.$$

Since the vectors $\underline{y}_1, \underline{y}_2, \cdots, \underline{y}_b$ are linearly independent with probability 1, (2.3.6) implies that

$$a_1 - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b} = 0 , \quad a_2 - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b} = 0 ,$$

$$a_{b-1} - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b} = 0 , \text{ and } - \frac{a_1 + a_2 + \cdots + a_{b-1}}{b} = 0 .$$

Hence $a_1 = a_2 = \cdots = a_{b-1} = 0$. But this is a contradiction, therefore rank $(Z) = b - 1$ with probability one.

Corollary 2.3.3 Let $KK' = I_b - \frac{1}{b} J_b$ such that K is $b \times b-1$ of rank $b-1$. If $t \geq b$ and if rank $(Y) = b$ with probability one, then rank $(K' Y' (I_t - \frac{1}{t} J_t) Y K) = b - 1$ with probability one.

Proof:

$$\text{rank } [K' Y' (I_t - \tfrac{1}{t} J_t) Y K] \geq \text{rank } [K K' Y' (I_t - \tfrac{1}{t} J_t) Y K K']$$

$$= b - 1 \text{ by Theorem 2.3.3.}$$

$$\text{rank } [K' Y' (I_t - \tfrac{1}{t} J_t) Y K] = \text{rank } [K' K K' Y' (I_t - \tfrac{1}{t} J_t) Y K K' K]$$

$$\leq \text{rank } [K K' Y' (I_t - \tfrac{1}{t} J_t) Y K K']$$

$$= b - 1 \text{ by Theorem 2.3.3}$$

and hence the result follows.

Theorem 2.3.4 Let Y be the $t \times b$ matrix of observations from a $t \times b$ two-way classification design with the model (2.1.1). Assume $t \geq b$. Then $K' Y' (I_t - \frac{1}{t} J_t) Y K$ is distributed as the Wishart $W_{b-1}(t-1, \sigma^2 I_{b-1}, K' \Gamma' (I_t - \frac{1}{t} J_t) \Gamma K)$ where K is defined in Corollary 2.3.3 and $\Gamma = (\gamma_{ij})$.

Proof: Let $U = Y(I_b - \frac{1}{b} J_b)$

$$= \begin{bmatrix} y_{11} - y_{1\cdot} & y_{12} - y_{1\cdot} & \cdots & y_{1b} - y_{1\cdot} \\ y_{21} - y_{2\cdot} & y_{22} - y_{2\cdot} & \cdots & y_{2b} - y_{2\cdot} \\ \vdots & \vdots & \cdot & \vdots \\ y_{t1} - y_{t\cdot} & y_{t2} - y_{t\cdot} & \cdots & y_{tb} - y_{t\cdot} \end{bmatrix}$$

Note that the rows of U are independent. Let

$U' = (I_b - \frac{1}{b} J_b) Y' = [\underline{u}_1 \underline{u}_2 \cdots \underline{u}_t]$. Hence $\underline{u}_i \sim$ independent

$N(\underline{\beta} + \underline{\gamma}_i, \sigma^2 (I_b - \frac{1}{b} J_b))$ $i = 1, 2, \cdots, t$ where

$\underline{\gamma}_i' = [\gamma_{i1} \gamma_{i2} \cdots \gamma_{ib}]$ $i = 1, 2, \cdots, t$. Now let $\underline{v}_i = K' \underline{u}_i$

$i = 1, 2, \cdots, t$, then $\underline{v}_i \sim$ independent $N(K'\underline{\beta} + K' \underline{v}_i, \sigma^2 I_{b-1})$

$i = 1, 2, \cdots, t$. Let $V' = [\underline{v}_1 \underline{v}_2 \cdots \underline{v}_t] = K'U' = K'KK'Y' = K'Y$

so that

$$K'Y'(I_t - \frac{1}{t} J_t) Y K = V'(I_t - \frac{1}{t} J_t) V .$$

Since $I_t - \frac{1}{t} J_t$ is idempotent of rank $t - 1$,

$$V'(I_t - \frac{1}{t} J_t) V \sim W_{b-1}(t - 1, \sigma^2 I_{b-1}, K' \Gamma'(I_t - \frac{1}{t} J_t) \Gamma K).$$

Therefore

(2.3.7)

$$K'Y'(I_t - \frac{1}{t} J_t) Y K \sim W_{b-1}(t - 1, \sigma^2 I_{b-1}, K' \Gamma'(I_t - \frac{1}{t} J_t) \Gamma K)$$

and the theorem is proved.

Corollary 2.3.4 The distribution of the $b - 1$ nonzero characteristic roots of $Z'Z$ is the same as the distribution of the characteristic roots of a matrix having the Wishart $W_{b-1}(t - 1, \sigma^2 I_{b-1}, K' \Gamma'(I_t - \frac{1}{t} J_t) \Gamma K)$ distribution.

Proof: By Corollary 2.3.2 the nonzero roots of $Z'Z$ are the same as the nonzero roots of $K'Y'(I_t - \frac{1}{t}J_t)YK$ and the result follows by Theorem 2.3.4.

The special case when $\xi(i, i', j, j') = 0$ for every i, i', j, j', which has previously been shown to be equivalent to $\Gamma = 0$, will now be considered. In this case the nonzero characteristic roots of $Z'Z$ are distributed the same as the characteristic roots of a matrix with the central Wishart $W_{b-1}(t-1, \sigma^2 I)$ distribution. The distribution of the roots of a central Wishart was found independently by Fisher (1939), Girshick (1939), Hsu (1939), Mood (1951), and Roy (1939). The distribution is

(2.3.8)

$$f(\ell_1, \ell_2, \cdots, \ell_{b-1})$$

$$= \frac{\pi^{\frac{b-1}{2}} \prod_{i=1}^{b-1} \ell_i^{\frac{t-b-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{b-1}\ell_i\right\} \prod_{i=1}^{b-2}\prod_{j=i+1}^{b-1}(\ell_i - \ell_j)}{(2\sigma^2)^{\frac{(b-1)(t-1)}{2}} \prod_{i=1}^{b-1}\left[\Gamma\left(\frac{t-i}{2}\right)\Gamma\left(\frac{b-i}{2}\right)\right]}$$

for $0 < \ell_{b-1} < \ell_{b-2} < \cdots < \ell_2 < \ell_1 < \infty$.

The distribution of the characteristic roots of a matrix with the non-central Wishart distribution was found by James (1961). The distribution depends only on σ^2 and the nonzero roots of the non-centrality matrix.

2.4 Determining Significance of Two by Two Table Differences

In Section 2.2 it was shown that

$$Q = \max_{i,i',j,j'} \frac{(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4}$$

$$\leq \ell_1 \leq \ell_1 + \ell_2 + \cdots + \ell_{b-1} = \mathrm{tr}(Z'Z) .$$

Therefore

$$\frac{Q}{\mathrm{tr}(Z'Z)} \leq \frac{\ell_1}{\mathrm{tr}(Z'Z)} \leq 1 .$$

Let

(2.4.1) $$u_1 = \frac{\ell_1}{\mathrm{tr}(Z'Z)} = \frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}} .$$

The distribution of $u_1 = \dfrac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}$ will now be found under the hypothesis Ho: $\xi(i,i',j,j') = 0$ for every $i, i', j,$ and j'.

From (2.3.8) the joint distribution of $\ell_1, \ell_2, \cdots, \ell_{b-1}$ under Ho is

(2.4.2)

$$f(\ell_1, \ell_2, \cdots, \ell_{b-1})$$

$$= C \prod_{i=1}^{b-1} \ell_i^{\frac{t-b-1}{2}} \exp\left\{\frac{1}{2\sigma^2} \sum_{i=1}^{b-1} \ell_i\right\} \prod_{i=1}^{b-2} \prod_{j=i+1}^{b-1} (\ell_i - \ell_j)$$

for $0 < \ell_{b-1} < \cdots < \ell_2 < \ell_1 < \infty$

where

$$(2.4.3) \qquad C = \frac{\pi^{\frac{b-1}{2}}}{(2\sigma^2)^{\frac{(b-1)(t-1)}{2}} \prod_{i=1}^{b-1} \left[\Gamma\left(\frac{t-i}{2}\right)\Gamma\left(\frac{b-i}{2}\right)\right]}$$

First make the following transformation. Let

$$u_1 = \frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}, \quad u_2 = \frac{\ell_2}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}, \cdots,$$

$$(2.4.4) \qquad u_{b-2} = \frac{\ell_{b-2}}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}, \quad u_{b-1} = \ell_1 + \ell_2 + \cdots + \ell_{b-1}.$$

Then

$$\ell_1 = u_1 u_{b-1}, \ \ell_2 = u_2 u_{b-1}, \cdots,$$

$$(2.4.5) \qquad \ell_{b-2} = u_{b-2} u_{b-1}, \ \ell_{b-1} = u_{b-1}(1 - u_1 - \cdots - u_{b-2}).$$

The jacobion of the transformation is

$$J = \left|\begin{array}{cccc:c} u_{b-1} & 0 & 0 \cdots & 0 & u_1 \\ 0 & u_{b-1} & 0 \cdots & 0 & u_2 \\ 0 & 0 & u_{b-1} \cdots & 0 & u_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & u_{b-1} & u_{b-2} \\ \hdashline -u_{b-1} & -u_{b-1} & -u_{b-1} \cdots & -u_{b-1} & 1 - u_1 - \cdots - u_{b-2} \end{array}\right| = \left|\begin{array}{cc} U_{11} & \underline{u}_{12} \\ \underline{u}'_{21} & u_{22} \end{array}\right| \text{ (say)}$$

$$(2.4.6) \qquad = |U_{11}|\,|u_{22} - \underline{u}_{12} U_{11}^{-1} \underline{u}_{12}|$$

by Theorem 8.2.1, Graybill (1969). Now

$$\underline{u}_{21} U_{11}^{-1} \underline{u}_{12} = (-u_{b-1} \underline{j}'_{b-2})(u_{b-1} I_{b-2})^{-1} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{b-2} \end{bmatrix}$$

$$= -(u_1 + u_2 + \cdots + u_{b-2}) .$$

Substituting this in (2.4.6), one obtains

$$J = |U_{11}| (1 - u_1 - u_2 - \cdots - u_{b-2} + u_1 + u_2 + \cdots + u_{b-2})$$

(2.4.7)

$$= u_{b-1}^{b-2} (1) = u_{b-1}^{b-2} .$$

Now

$$0 < \ell_1 + \ell_2 + \cdots + \ell_{b-1} < \infty \text{ implies } 0 < u_{b-1} < \infty$$

and

$$\ell_1 + \ell_2 + \cdots + \ell_{b-1} > \ell_1 > \ell_2 > \cdots > \ell_{b-1} > 0$$

implies

$$u_{b-1} > u_1 u_{b-1} > u_2 u_{b-1} > \cdots > u_{b-2} u_{b-1}$$

$$> u_{b-1} (1 - u_1 - u_2 - \cdots - u_{b-2}) > 0$$

which implies

$$1 > u_1 > u_2 > \cdots > u_{b-2} > 1 - u_1 - u_2 - \cdots - u_{b-2} > 0.$$

Equivalently one can write the following :

$$\frac{1-u_1-\cdots-u_{b-3}}{2} < u_{b-2} < \min(u_{b-3}, 1-u_1-\cdots-u_{b-1})$$

$$\frac{1-u_1-\cdots-u_{b-4}}{3} < u_{b-3} < \min(u_{b-4}, 1-u_1-\cdots-u_{b-4})$$

(2.4.8) $$\vdots$$

$$\frac{1-u_1}{b-2} < u_2 < \min(u_1, 1-u_1)$$

$$\frac{1}{b-1} < u_1 < 1$$

and

(2.4.9) $$0 < u_{b-1} < \infty \quad .$$

Therefore the joint density of $u_1, u_2, \cdots, u_{b-1}$ is given by

$$g(u_1, u_2, \cdots, u_{b-1})$$

$$= C \prod_{i=1}^{b-2} [u_i u_{b-1}]^{\frac{t-b-1}{2}} [u_{b-1}(1-u_1-\cdots-u_{b-2})]^{\frac{t-b-1}{2}}$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2} u_{b-1}\right\} \prod_{i=1}^{b-3} \prod_{j=i+1}^{b-2} (u_{b-1}u_i - u_{b-1}u_j)$$

$$\cdot \prod_{i=1}^{b-2} [u_i u_{b-1} - u_{b-1}(1-u_1-\cdots-u_{b-2})] \cdot u_{b-1}^{b-2}$$

$$= C\left(\prod_{i=1}^{b-2} u_i^{\frac{t-b-1}{2}}\right)(1-u_1-\cdots-u_{b-2})^{\frac{t-b-1}{2}} \exp\left\{-\frac{1}{2\sigma^2} u_{b-1}\right\}$$

$$\cdot \prod_{i=1}^{b-3} \prod_{j=i+1}^{b-2} (u_i - u_j) \prod_{i=1}^{b-2} (u_i + u_1 + u_2 + \cdots + u_{b-2} - 1) \cdot u_{b-1}^{\frac{(b-1)(t-1)}{2} - 1}$$

$$= \pi^{\frac{b-1}{2}} \left[(1 - u_1 - \cdots - u_{b-2}) \prod_{i=1}^{b-2} u_i \right]^{\frac{t-b-1}{2}} \left[\prod_{i=1}^{b-3} \prod_{j=i+1}^{b-2} (u_i - u_j) \right]$$

(2.4.10)

$$\prod_{i=1}^{b-2} (u_i - 1 + u_1 + \cdots + u_{b-2}) \Big/ \prod_{i=1}^{b-1} \left[\Gamma\left(\frac{t-i}{2}\right) \Gamma\left(\frac{b-i}{2}\right) \right]$$

$$\cdot \frac{1}{(2\sigma^2)^{\frac{(b-1)(t-1)}{2}}} u_{b-1}^{\frac{(b-1)(t-1)}{2} - 1} e^{-\frac{u_{b-1}}{2\sigma^2}}$$

for range given in (2.4.8) and (2.4.9). From (2.4.10) one can see that u_{b-1} is independent of $u_1, u_2, \cdots, u_{b-2}$ since the density (2.4.10) can be factored. Notice that

$$\int_0^\infty \frac{u_{b-1}^{\frac{(b-1)(t-1)}{2} - 1} e^{-\frac{u_{b-1}}{2\sigma^2}}}{(2\sigma^2)^{\frac{(b-1)(t-1)}{2}}} = \Gamma\left(\frac{(b-1)(t-1)}{2}\right), \tag{2.4.11}$$

thus the joint density of $u_1, u_2, \cdots, u_{b-2}$ under Ho is

(2.4.12)

$$g^*(u_1, u_2, \cdots, u_{b-2})$$

$$= C^* \left[u_1 u_2 \cdots u_{b-2} (1 - u_1 - u_2 - \cdots - u_{b-2}) \right]^{\frac{t-b-1}{2}}$$

$$\prod_{i=1}^{b-3} \prod_{j=i+1}^{b-2} (u_i - u_j) \prod_{i=1}^{b-2} (u_i - 1 + u_1 + u_2 + \cdots + u_{b-2})$$

for range given in (2.4.8) where

(2.4.13)
$$C^* = \frac{\pi^{\frac{b-1}{2}} \Gamma\left(\frac{(b-1)(t-1)}{2}\right)}{\prod_{i=1}^{b-1} \left[\Gamma\left(\frac{t-i}{2}\right) \Gamma\left(\frac{b-i}{2}\right)\right]}.$$

Notice that the density in (2.4.12) is independent of all parameters. The marginal density of u_1 under Ho is then given by

$$g_1(u_1) = \int_{\frac{1-u_1}{b-2}}^{a_1} \int_{\frac{1-u_1-u_2}{b-3}}^{a_2} \cdots \int_{\frac{1-u_1-\cdots-u_{b-3}}{2}}^{a_{b-3}}$$

(2.4.14)

$$g^*(u_1, u_2, \cdots, u_{b-2})\, du_{b-2} \cdots du_3\, du_2$$

for $\frac{1}{b-1} < u_1 < 1$ where

$$a_1 = \min(u_1, 1-u_1),\ a_2 = \min(u_2, 1-u_1-u_2), \cdots,$$

$$a_{b-3} = \min(u_{b-3}, 1-u_1-\cdots-u_{b-3}).$$

This density will be simplified in Chapter 4 for the special cases $b = 3$ and $b = 4$.

Now a critical point will be obtained for determining which (if any) of the estimates of two by two table differences are significantly different from zero.

Theorem 2.4.1 Let $0 < \alpha < 1$ be given. Let X_α be the number satisfying (i) $\frac{1}{b-1} < X_\alpha < 1$ and (ii) $\int_{1/b-1}^{X_\alpha} g_1(u_1)\,du_1 = 1 - \alpha$.

Let $Q = \max_{i, i', j, j'} \frac{(y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'})^2}{4}$. Then

$$P_{Ho}\left\{ Q \leq \frac{\ell_2 + \cdots + \ell_{b-1}}{1 - X_\alpha} \cdot X_\alpha \right\} \geq 1 - \alpha .$$

Proof:

$$P_{Ho}\left\{ u_1 \leq X_\alpha \right\} = \int_{1/b-1}^{X_\alpha} g_1(u_1)\,du_1 = 1 - \alpha$$

implies

$$P_{Ho}\left\{ \frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}} \leq X_\alpha \right\} = 1 - \alpha$$

which implies

$$(2.4.15) \qquad P_{Ho}\left\{ \ell_1(1 - X_\alpha) \leq (\ell_2 + \cdots + \ell_{b-1}) X_\alpha \right\} = 1 - \alpha .$$

Since $Q \leq \ell_1$, (2.4.15) implies that

$$P_{Ho}\left\{ Q(1 - X_\alpha) \leq (\ell_2 + \cdots + \ell_{b-1}) X_\alpha \right\} \geq 1 - \alpha .$$

Hence

$$P_{Ho}\left\{ Q \leq \frac{X_\alpha(\ell_2 + \cdots + \ell_{b-1})}{1 - X_\alpha} \right\} \geq 1 - \alpha .$$

Corollary 2.4.1 For α and X_α defined in Theorem 2.4.1,

$$P_{Ho}\left\{ |y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'}| \leq 2\sqrt{\frac{X_\alpha(\ell_2 + \cdots + \ell_{b-1})}{1 - X_\alpha}} \text{ for all } i, i', j, j' \right\}$$

$$\geq 1 - \alpha .$$

Proof: Let $\hat{\xi}(i, i', j, j') = y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'}$.

$$P_{Ho}\left\{ Q \leq \frac{(\ell_2 + \cdots + \ell_{b-1}) X_\alpha}{1 - X_\alpha} \right\} \geq 1 - \alpha \qquad \text{if and only if}$$

$$P_{Ho}\left\{ [\hat{\xi}(i, i', j, j')]^2/4 \leq \frac{X_\alpha(\ell_2 + \cdots + \ell_{b-1})}{1 - X_\alpha} \text{ for all } i, i', j, j' \right\}$$

$$\geq 1 - \alpha .$$

Hence

$$P_{Ho}\left\{ |\hat{\xi}(i, i', j, j')| \leq 2\sqrt{\frac{X_\alpha(\ell_2 + \cdots + \ell_{b-1})}{1 - X_\alpha}} \text{ for all } i, i', j, j' \right\}$$

$$\geq 1 - \alpha .$$

By Corollary 2.4.1 $2\sqrt{\frac{X_\alpha(\ell_2 + \cdots + \ell_{b-1})}{1 - X_\alpha}}$ is a critical point for determining which (if any) of the estimates of two by two table differences are significantly different from zero.

In the joint density of $u_1, u_2, \cdots, u_{b-2}$ given by (2.4.12), observe that the distribution of each of the statistics $u_1, u_2, \cdots, u_{b-2}$ does not depend on the values of σ^2; that is, each of the statistics

$$\frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}, \frac{\ell_2}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}, \cdots,$$

$$\frac{\ell_{b-2}}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}, \frac{\ell_{b-1}}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}} = 1 - u_1 - u_2 - \cdots - u_{b-2}$$

are independent of the value of σ^2. Also note that

$$\frac{\ell_1}{\ell_2 + \ell_3 + \cdots + \ell_{b-1}} = \frac{u_1}{1 - u_1}, \frac{\ell_1}{\ell_3 + \ell_4 + \cdots + \ell_{b-1}} = \frac{u_1}{1 - u_1 - u_2},$$

$$\cdots, \frac{\ell_1}{\ell_{b-1}} = \frac{u_1}{1 - u_1 - \cdots - u_{b-2}}$$

are each independent of σ^2. Therefore the distribution of any of the statistics above could be used to find a critical point for determining which of the two by two table differences are significantly different from zero. In this paper only the statistic

$\frac{\ell_1}{\ell_2 + \cdots + \ell_{b-1}}$ which is equivalent to $\frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}$

will be considered.

2.5 Estimates of σ^2 and Hypothesis Testing

Let the critical point for determining the significance of estimates of two by two table difference found in Corollary 2.4.1 be represented by Z_α; that is, let

$$Z_\alpha = 2\sqrt{\frac{X_\alpha(\ell_2 + \ell_3 + \cdots + \ell_{b-1})}{1 - X_\alpha}} \quad . \tag{2.5.1}$$

Let Z_α^* be the point that satisfies

$$P_{Ho}\left\{|y_{ij} - y_{i'j} - y_{ij'} + y_{i'j'}| < Z_\alpha^* \text{ for all } i, i', j, j'\right\} = 1 - \alpha \quad . \tag{2.5.2}$$

Thus one has

$$Z_\alpha^* \leq Z_\alpha \quad . \tag{2.5.3}$$

For the time being it will be assumed that Z_α^* is known.

Using the critical point Z_α^* defined in (2.5.2), one can divide the $\frac{bt(b-1)(t-1)}{4}$ distinct two by two tables into two sets. The first set will contain those two by two tables which have estimated differences less than Z_α^*; that is, those that are not significantly different from zero. The second set will contain those two by two table differences which have estimated differences which are greater than Z_α^*; that is, those that are significantly different from zero. Suppose that the first set contains m distinct two by two tables and the second set contains n distinct two by two tables so that $m + n = \frac{bt(b-1)(t-1)}{4}$. Let A be a $m \times bt$ matrix containing

those contrasts as rows which designate two by two table differences corresponding to the two by two tables in the first set. Let B be a $n \times bt$ matrix containing those contrasts as rows which designate two by two table differences corresponding to those two by two tables in the second set. Thus A contains contrasts of the observations which are not significantly different from zero and B contains contrasts of the observations which are significantly different from zero.

Ideally one would like none of the contrasts in the row space of A which designate two by two table differences to be a row of the matrix B. For example, if z_1 and z_2 represent two by two tables differences which are not significantly different from zero and if $z_1 - z_2 = z_3$ also represents a two by two table difference, one would hope that z_3 is not significantly different from zero. Unfortunately this is not always the case. For this reason two estimates of σ^2 will be proposed, the first in Corollary 2.5.1 and the second in Corollary 2.5.2.

Let $\xi_1, \xi_2, \cdots, \xi_m$ be those two by two table differences whose estimates $z_1, z_2, \cdots, z_m$ are not significantly different from zero. For example ξ_1 might be $\mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}$ and z_1 would then be $y_{11} - y_{12} - y_{21} + y_{22}$. Thus A is the matrix satisfying $A\underline{\mu} = \underline{\xi}$ where $\underline{\mu}' = [\mu_{11}\mu_{12}\cdots\mu_{1b}\mu_{21}\mu_{22}\cdots\mu_{2b}\cdots\mu_{t1}\mu_{t2}\cdots\mu_{tb}]$ and $\underline{\xi}' = [\xi_1\xi_2\cdots\xi_m]$. Similarly, let $\xi_1^*\xi_2^*\cdots\xi_n^*$ be those two by two table differences whose estimates $z_1^*z_2^*\cdots z_n^*$ are significantly different from zero. Thus B is the matrix satisfying $B\underline{\mu} = \underline{\xi}^*$ where $\underline{\mu}$ is the same as defined above and $\underline{\xi}^{*\prime} = [\xi_1^*\xi_2^*\cdots\xi_n^*]$.

Theorem 2.5.1 Let A be the matrix defined above and let $\underline{y}$ be the vector of observations from the two-way classification model (2.1.1).

If $\underline{\xi} = \underline{0}$, then $\underline{y}' \frac{A^- A}{\sigma^2} \underline{y}$ has the chi-square distribution with k degrees of freedom where k equals the rank of A.

Proof: $A^- A$ is symmetric idempotent by definition so one needs only to show that

$$\lambda = \frac{1}{2\sigma^2} \underline{\mu}' A^- A \underline{\mu} = \underline{0} .$$

Now $A\underline{\mu} = \underline{\xi} = 0$ by hypothesis, hence $\lambda = 0$ and the result follows by Corollary 4.7.1, Graybill (1961).

Corollary 2.5.1 If $\underline{\xi} = \underline{0}$, then $\hat{\sigma}^2_A = \frac{\underline{y}' A^- A}{k} \underline{y}$ is a k degree of freedom estimate of σ^2 where A, $\underline{y}$, and k are defined as in Theorem 2.5.1.

Proof: The result follows upon application of Theorem 2.5.1 and Definition 2.1.2.

$\hat{\sigma}^2_A$ will be called a conditional k degree of freedom estimate of σ^2. It is called conditional since the proof of Theorem 2.5.1 required that $\underline{\xi} = \underline{0}$; that is, it is assumed that the two by two table differences that were put in the first set are truly estimates of zero.

Theorem 2.5.2 Let A and B be defined as above and let $\underline{y}$ be defined as in Theorem 2.5.1. If $\underline{\xi} = \underline{0}$, then

$$\frac{\underline{y}' \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - B^- B \right] \underline{y}}{\sigma^2}$$

has the chi-square distribution with $(b - 1)(t - 1) - p$ degrees of freedom where p equals the rank of B.

Proof: First note that

$$\left[\left(I_b - \frac{1}{b} J_b\right) \otimes \left(I_t - \frac{1}{t} J_t\right)\right] B'$$

(2.5.4)

$$= \left(I_b \otimes I_t - \frac{1}{b} J_b \otimes I_t - \frac{1}{t} J_b \otimes J_t + \frac{1}{bt} J_b \otimes J_t\right) B'$$

$$= \left(I_b \otimes I_t\right) B' = B' .$$

Thus it is easily shown that $\left(I_b - \frac{1}{b} J_b\right) \otimes \left(I_t - \frac{1}{t} J_t\right) - B^- B$ is symmetric idempotent and hence the

$$\text{rank}\left[\left(I_b - \frac{1}{b} J_b\right) \otimes \left(I_t - \frac{1}{t} J_t\right) - B^- B\right]$$

$$= \text{tr}\left[\left(I_b - \frac{1}{b} J_b\right) \otimes \left(I_t - \frac{1}{t} J_t\right) - B^- B\right]$$

$$= (b-1)(t-1) - p \quad .$$

Therefore one needs only show that

(2.5.5) $$\lambda = \frac{1}{2\sigma^2} \underline{\mu}' \left[\left(I_b - \frac{1}{b} J_b\right) \otimes \left(I_t - \frac{1}{t} J_t\right) - B^- B\right] \underline{\mu} = 0 .$$

Let C be the $(m + n) \times bt$ matrix containing all the contrasts as rows which designate two by two table differences as rows. It is then clear that $\text{Col}(B') \subset \text{Col}(C')$ which is equivalent to $\text{Col}(B'B'^-) \subset \text{Col}(C'C'^-)$ which is equivalent to $\text{Col}(B^- B) \subset \text{Col}(C^- C)$. The intersection of the $\text{Col}(C^- C)$ and the orthogonal complement of the $\text{Col}(B^- B)$ is the same as the $\text{Col}(C^- C - B^- B)$ and it is clear that the $\text{Col}(C^- C - B^- B) \subset \text{Col}(A)$. Therefore $\underline{\xi} = \underline{0}$ implies $A\underline{\mu} = \underline{0}$, which implies that $(C^- C - B^- B)\underline{\mu} = \underline{0}$. Then equation (2.5.5) follows since $\text{Col}(C^- C) = \text{Col}\left(\left(I_b - \frac{1}{b} J_b\right) \otimes \left(I_t - \frac{1}{t} J_t\right)\right)$. Hence by Corollary 4.7.1, Graybill (1961) the theorem follows.

Corollary 2.5.2 If $\underline{\xi} = \underline{0}$, then

$$\hat{\sigma}_B^2 = \frac{\underline{y}' \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - B^- B \right] \underline{y}}{(b-1)(t-1) - p}$$

is a $(b-1)(t-1) - p$ degree of freedom estimate of σ^2 where $\underline{\xi}$, B, $\underline{y}$, and p are defined in Theorem 2.5.2.

Proof: The result follows upon application of Theorem 2.5.2 and Definition 2.1.2.

$\hat{\sigma}_B^2$ will be called a conditional $(b-1)(t-1) - p$ degree of freedom estimate of σ^2.

Corollary 2.5.3 If $\underline{\xi} = \underline{0}$, then $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ are unbiased estimates of σ^2.

Proof: By Theorem 2.5.1 and Theorem 2.5.2,

$$\frac{k\,\hat{\sigma}_A^2}{\sigma^2} \sim \chi^2(k) \quad \text{and} \quad \frac{\left[(b-1)(t-1) - p\right]\hat{\sigma}_B^2}{\sigma^2} \sim \chi^2\left((b-1)(t-1) - p\right).$$

Since the expected value of a chi-square random variable is equal to its degrees of freedom, the results follow.

Corollary 2.5.4 If $\underline{\xi} = \underline{0}$, the intervals

$$(2.5.6) \qquad \left(k\,\hat{\sigma}_A^2 \Big/ \chi^2_{\frac{\alpha}{2}}(k),\ k\hat{\sigma}_A^2 \Big/ \chi^2_{1-\frac{\alpha}{2}}(k) \right)$$

and

$$(2.5.7) \qquad \left(q\,\sigma_B^2 \Big/ \chi^2_{\frac{\alpha}{2}}(q),\ q\,\sigma_B^2 \Big/ \chi^2_{1-\frac{\alpha}{2}}(q) \right),$$

where $q = (b-\alpha)(t-\alpha) - p$

and $\chi^2_\alpha(n)$ is the point satisfying

$$\int_0^{\chi^2_\alpha(n)} \frac{u^{n-1} e^{-u/2}}{2^n \Gamma(n)} \, du = 1 - \alpha \text{ , are } (1-\alpha)\ 100\%$$

confidence intervals for σ^2.

Proof: Since $\frac{k\hat{\sigma}^2_A}{\sigma^2} \sim \chi^2(k)$,

$$P\left\{ \chi^2_{1-\frac{\alpha}{2}}(k) < \frac{k\hat{\sigma}_A}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(k) \right\} = 1 - \alpha$$

and since $\frac{q\hat{\sigma}^2_B}{\sigma^2} \sim \chi^2(q)$,

$$P\left\{ \chi^2_{1-\frac{\alpha}{2}}(q) < \frac{q\hat{\sigma}^2_B}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(q) \right\} = 1 - \alpha \text{ ,}$$

from which the results follow.

Two estimates of σ^2 have been proposed. Some of the advantages and disadvantages of each will now be discussed. In the proof of Theorem 2.5.2 it was stated that the $\text{Col}(C^- C - B^- B) \subset \text{Col}(A^- A)$ which implies that the $\text{rank}(C^- C - B^- B) \leq \text{rank}(A^- A)$ which implies that $(b-1)(t-1) - p \leq k$. Thus it is seen that the estimate $\hat{\sigma}^2_A$ is always based on at least as many degrees of freedom as the estimate $\hat{\sigma}^2_B$. For this reason one may prefer the estimate $\hat{\sigma}^2_A$ to the estimate $\hat{\sigma}^2_B$. Earlier in this section it was stated that one would like none of the vectors in the row space of A to be rows of the matrix B. If this is not the case then there exists at least one linear combination

of the rows of A which is a row of B. That is, the Col $(A^- A)$ will contain at least one two by two table difference which was significantly different from zero; so that, the estimate $\hat{\sigma}_A^2$ is actually based on some estimated two by two table differences which were found to be significantly different from zero. On the other hand the estimate $\hat{\sigma}_B^2$ is based only on estimated two by two table differences which were not significantly different from zero. Finally, although the matrix B contains as rows only those contrasts designating two by two tables which are significantly different from zero, it seems very likely that in some cases many of the rows of A could be included in the row space of B and thus would not be used in the estimate $\hat{\sigma}_B^2$. It is the opinion of this writer that this is the most important factor of the three. For this reason the estimate $\hat{\sigma}_A^2$ is recommended rather than the estimate $\hat{\sigma}_B^2$ unless a large percentage of the vectors in the row space of A are rows of the matrix B.

The following two theorems give statistics which have the noncentral F distribution with noncentrality parameter $\lambda = \frac{1}{2\sigma^2} \sum_{i=1}^{t} \sum_{j=1}^{b} \gamma_{ij}^2$. Thus these statistics could be used to give some indication of how much non-additivity has actually been accounted for. If the statistic is close to 1, one could not have nearly the faith in the estimates of σ^2 as one would have if the statistic is quite large.

Theorem 2.5.3 Let A, k, and ξ be defined as above. If $\underline{\xi} = 0$, then

$$(2.5.8) \qquad \frac{\underline{y}' \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - A^- A \right] \underline{y}}{\underline{y}' A^- A \underline{y}} \quad \frac{k}{(b-1)(t-1) - k}$$

has the noncentral F distribution with $(b-1)(t-1) - k$ and k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2\sigma^2} \sum_{i=1}^{t} \sum_{j=1}^{b} \gamma_{ij}^2$.

Proof: By Theorem 2.5.1 $\frac{\underline{y}' A^- A \underline{y}}{\sigma^2} \sim \chi^2(k)$ if $\underline{\xi} = \underline{0}$.

$$(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) A' = A' ,$$

hence

$$U = (I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - A^- A$$

is symmetric idempotent and rank (U) = tr (U) = (b - 1)(t - 1) - k.

Therefore $\frac{\underline{y}' U \underline{y}}{\sigma^2} \sim \chi'^2 ((b-1)(t-1) - k, \lambda)$ by Corollary 4.7.1, Graybill (1961) where

$$\lambda = \frac{1}{2\sigma^2} \underline{\mu}' \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - A^- A \right] \underline{\mu}$$

$$= \frac{1}{2\sigma^2} \underline{\mu}' \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) \right] \underline{\mu}$$

since $A\underline{\mu} = \underline{\xi} = \underline{0}$. Thus

$$\lambda = \frac{1}{2\sigma^2} \sum_{i=1}^{t} \sum_{j=1}^{b} \gamma_{ij}^2 .$$

Now

$$A^- A \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - A^- A \right] = A^- A - A^- A = 0 ,$$

so that $\frac{\underline{y}' A^- A \underline{y}}{k}$ is independent of

$$\frac{\underline{y}' \left[(I_b - \frac{1}{b} J_b) \otimes (I_t - \frac{1}{t} J_t) - A^- A \right] \underline{y}}{(b-1)(t-1) - k} .$$

Therefore

$$\frac{\underline{y}' \left[(I_b - \frac{1}{b}J_b) \otimes (I_t - \frac{1}{t}J_t) - A^- A \right] \underline{y}}{\underline{y}' A^- A \underline{y}} \cdot \frac{k}{(b-1)(t-1)-k}$$

$$\sim F'((b-1)(t-1)-k,\ k,\ \lambda) .$$

Theorem 2.5.4 Let A, B, p, and ξ be defined as above. If $\underline{\xi} = \underline{0}$, then

$$(2.5.9) \qquad \frac{\underline{y}' B^- B \underline{y}}{\underline{y}' \left[(I_b - \frac{1}{b}J_b) \otimes (I_t - \frac{1}{t}J_t) - B^- B \right] \underline{y}} \cdot \frac{(b-1)(t-1)-p}{p}$$

has the noncentral F distribution with p and $(b-1)(t-1)-p$ degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2\sigma^2} \sum_{i=1}^{t} \sum_{j=1}^{b} \gamma_{ij}^2$.

Proof: By Theorem 2.5.2

$$\frac{\underline{y}' \left[(I_b - \frac{1}{b}J_b) \otimes (I_t - \frac{1}{t}J_t) - B^- B \right] \underline{y}}{\sigma^2} \sim \chi^2((b-1)(t-1)-p)$$

if $\underline{\xi} = \underline{0}$. $B^- B$ is symmetric idempotent of rank p, hence $\frac{\underline{y}' B^- B \underline{y}}{\sigma^2} \sim \chi'^2(p, \lambda)$ where

$$\lambda = \frac{1}{2\sigma^2} \underline{\mu}' B^- B \underline{\mu}$$

$$= \frac{1}{2\sigma^2} \underline{\mu}' \left[(I_b - \frac{1}{b}J_b) \otimes (I_t - \frac{1}{t}J_t) - \left[(I_b - \frac{1}{b}J_b) \otimes (I_t - \frac{1}{t}J_t) - B^- B \right] \right] \underline{\mu}$$

$$= \frac{1}{2\sigma^2} \underline{\mu}' \left[\left(I_b - \frac{1}{b} J_b \right) \otimes \left(I_t - \frac{1}{t} J_t \right) \right] \underline{\mu}$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{t} \sum_{j=1}^{b} \gamma_{ij}^2 .$$

Finally,

$$B^- B \left[\left(I_b - \frac{1}{b} J_b \right) \otimes \left(I_t - \frac{1}{t} J_t \right) - B^- B \right] = B^- B - B^- B = 0 ,$$

hence $\frac{\underline{y}' B^- B \underline{y}}{p}$ is independent of

$$\frac{\underline{y}' \left[\left(I_b - \frac{1}{b} J_b \right) \otimes \left(I_t - \frac{1}{t} J_t \right) - B^- B \right] \underline{y}}{(b-1)(t-1) - p} .$$

Therefore

$$\frac{\underline{y}' B^- B \underline{y}}{\underline{y}' \left[\left(I_b - \frac{1}{b} J_b \right) \otimes \left(I_t - \frac{1}{t} J_t \right) - B^- B \right] \underline{y}} \cdot \frac{(b-1)(t-1) - p}{p}$$

$$\sim F'(p, (b-1)(t-1) - p, \lambda) .$$

One can also construct conditional tests of hypothesis concerning the treatment effects and set conditional confidence intervals on estimable contrasts of the treatment effects. To do this one needs only to use one of the estimates of σ^2 given here with its degrees of freedom in any of the available tests or confidence intervals requiring an independent estimate of σ^2.

At the beginning of this section it was assumed that Z_α^* was known. At this time Z_α^* is not known but $Z_\alpha \geq Z_\alpha^*$ can be found. If one uses Z_α rather than Z_α^* as the point to determine which contrasts belong to the matrix A and which belong to the matrix B,

then one concludes more of the two by two table differences estimate zero. Thus one is actually operating with the probability of Type I error less than what is indicated by α . In other words, the estimate one gets as an estimate of σ^2 will very likely be conservative; i.e., an overestimate of σ^2.

In the next section an example is given to illustrate the concepts discussed here.

2.6 An Example

In order to illustrate the topics discussed in Section 2.5, consider the following two-way classification design and data from Black (1970). The observations are yields in kg/ha of a spring wheat as affected by various combinations of levels of nitrogen and phosphorus.

Table 2.6.1 Yield in kg/ha of spring wheat

Nitrogen in kg/ha	Phosphorus in kg/ha					Totals
	0	22	45	90	180	
0	1984	2550	2706	2740	2954	12934
45	1776	2843	3306	3305	3386	14616
90	1797	2761	3240	3227	3332	14357
Totals	5557	8154	9252	9272	9672	

The analysis of variance of the data in Table 2.6.1 is given in Table 2.6.2 where the single degree of freedom for non-additivity is that proposed by Tukey (1949).

Table 2.6.2 Yield in kg/ha of spring wheat, Analysis of Variance

Source	df	S S	M S	
Total	14	4,335,296.40		
Nitrogen	2	328,075.60		
Phosphorus	4	3,748,569.07		
Non-additivity	1	235,766.56	235,766.56	72.11
Residual	7	22,885.17	3,269.31	

The probability of attaining a value as large as 72.11, if there is no interaction in the data, is less than .0001. Since one must conclude that there is interaction in the data, the estimate of σ^2 which has been available for use is 258,651.73/8 which is equal to 32,331.47.

In order to find the estimate of σ^2 proposed in Section 2.5, one needs from (2.5.1)

$$Z_\alpha = 2\sqrt{\frac{X_\alpha \ell_2}{1 - X_\alpha}}$$

where ℓ_2 is the smallest characteristic root of $Z'Z$ and X_α is defined in Theorem 2.4.1. From the data one obtains $\ell_1 = 257,721.16$ and $\ell_2 = 930.57$. From Table 4.4.1, $X_{.05} = .92967$. Thus

$$(2.6.1) \qquad Z_{.05} = 2\sqrt{\frac{(.92967)(930.57)}{.07033}} = 221.8189 .$$

The two by two table differences of the data in Table 2.6.1 are given in Table 2.6.3.

Table 2.6.3 Two by Two Table Differences

i i' j j'	$\hat{\xi}(i,i',j,j')$	i i' j j'	$\hat{\xi}(i,i',j,j')$	i i' j j'	$\hat{\xi}(i,i',j,j')$
1 2 1 2	501*	1 3 1 2	398*	2 3 1 2	-103
1 2 1 3	808*	1 3 1 3	721*	2 3 1 3	-87
1 2 1 4	773*	1 3 1 4	674*	2 3 1 4	-99
1 2 1 5	640*	1 3 1 5	565*	2 3 1 5	-75
1 2 2 3	307*	1 3 2 3	323*	2 3 2 3	16
1 2 2 4	272*	1 3 2 4	276*	2 3 2 4	4
1 2 2 5	139	1 3 2 5	167	2 3 2 5	28
1 2 3 4	-35	1 3 3 4	-47	2 3 3 4	-12
1 2 3 5	-168	1 3 3 5	-156	2 3 3 5	12
1 2 4 5	-133	1 3 4 5	-109	2 3 4 5	24

*Significantly different from zero .

The matrix A defined in Section 2.5 is given by (2.6.2)

$$A = \left[\begin{array}{rrrrr:rrrrr:rrrrr}
0 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1
\end{array}\right]$$

and $A^{-}A$ is then given by

(2.6.3) $120\,A^{-}A =$

$$\left[\begin{array}{rrrrr:rrrrr:rrrrr}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 60 & -20 & -20 & -20 & 0 & -30 & 10 & 10 & 10 & 0 & -30 & 10 & 10 & 10 \\
0 & -20 & 60 & -20 & -20 & 0 & 10 & -30 & 10 & 10 & 0 & 10 & -30 & 10 & 10 \\
0 & -20 & -20 & 60 & -20 & 0 & 10 & 10 & -30 & 10 & 0 & 10 & 10 & -30 & 10 \\
0 & -20 & -20 & -20 & 60 & 0 & 10 & 10 & 10 & -30 & 0 & 10 & 10 & 10 & -30 \\
\hdashline
0 & 0 & 0 & 0 & 0 & 48 & -12 & -12 & -12 & -12 & -48 & 12 & 12 & 12 & 12 \\
0 & -30 & 10 & 10 & 10 & -12 & 63 & -17 & -17 & -17 & 12 & -33 & 7 & 7 & 7 \\
0 & 10 & -30 & 10 & 10 & -12 & -17 & 63 & -17 & -17 & 12 & 7 & -33 & 7 & 7 \\
0 & 10 & 10 & -30 & 10 & -12 & -17 & -17 & 63 & -17 & 12 & 7 & 7 & -33 & 7 \\
0 & 10 & 10 & 10 & -30 & -12 & -17 & -17 & -17 & 63 & 12 & 7 & 7 & 7 & -33 \\
\hdashline
0 & 0 & 0 & 0 & 0 & -48 & 12 & 12 & 12 & 12 & 48 & -12 & -12 & -12 & -12 \\
0 & -30 & 10 & 10 & 10 & 12 & -33 & 7 & 7 & 7 & -12 & 63 & -17 & -17 & -17 \\
0 & 10 & -30 & 10 & 10 & 12 & 7 & -33 & 7 & 7 & -12 & -17 & 63 & -17 & -17 \\
0 & 10 & 10 & -30 & 10 & 12 & 7 & 7 & -33 & 7 & -12 & -17 & -17 & 63 & -17 \\
0 & 10 & 10 & 10 & -30 & 12 & 7 & 7 & 7 & -33 & -12 & -17 & -17 & -17 & 63
\end{array}\right]$$

and hence

$$(A^{-}A\underline{y})' = \frac{1}{120}\,[0 \;\; 14840 \;\; -10360 \;\; -7080 \;\; 2600 \;\; -4368 \;\; -5608 \;\; 6032$$

$$5112 \;\; -1168 \;\; 4368 \;\; -9232 \;\; 4328 \;\; 1968 \;\; -1432\,].$$

Therefore

(2.6.4)

$$\hat{\sigma}^2 = \frac{\underline{y}'A^{-}A\underline{y}}{\text{rank}\,(A)} = \frac{(\underline{y}'A^{-}A)(A^{-}A\underline{y})}{\text{tr}\,(A^{-}A)} = \frac{43598.4}{7} = 6228.34286\,.$$

Observe that this is much smaller than the estimate obtained without taking into account the non-additivity which was 32,331.47.

Careful study of Table 2.6.3 reveals some interesting aspects of the pattern of non-additivity or the nature of the interaction. First notice that nitrogen levels 2 and 3 appear to be free of interaction while significant differences occur when both nitrogen levels 2 and 3 occur with level 1 . Also notice that when nitrogen levels 2 and 3 occur with level 1 there are no significant differences among phosphorus levels 3,4 and 5 . Thus it seems that all of the non-additivity present in the data is caused by phosphorus levels 1 and 2

in the presence of nitrogen level 1 . The conditional analysis of variance as discussed in Section 2.5 is given in Table 2.6.4 .

Table 2.6.4 Conditional AoV

SV	df	SS	MS	F
Total	14	4,335,296.40		
Nitrogen	2	328,075.60	164,037.80	26.34
Phosphorus	4	3,748,569.07	937,142.27	150.46
Interaction	1	215,053.33	215,053.33	34.53
Error	7	43,598.40	6,228.34	

One may now do anything in the way of analysis of the data that might ordinarily be done using the error mean square above as an estimate of σ^2 . However, keep in mind the fact that the results are conditional. Remember also that since it seems that the estimate of σ^2 is an overestimate, one is still quite safe as far as the significance level is concerned in the statement made.

CHAPTER 3

THE MODEL $y_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\gamma_j + e_{ij}$

3.1 Introduction

In the model (2.1.1) the interaction term γ_{ij} will be defined by

$$(3.1.1) \qquad \gamma_{ij} = \lambda\alpha_i\gamma_j \quad i = 1, 2, \cdots, t; \; j = 1, 2, \cdots, b \quad (t \geq b),$$

where $\alpha_1, \alpha_2, \cdots, \alpha_t$ is one set of unknown parameters defined to be the block contribution to the interaction effect, and λ is an unknown scale parameter defined to be the magnitude of the interaction effect. In addition to the assumption that

$$(3.1.2) \qquad \sum_{i=1}^{t} \tau_i = \sum_{j=1}^{b} \beta_j = 0 \quad ,$$

it will be assumed that

$$(3.1.3) \qquad \sum_{i=1}^{t} \alpha_i = \sum_{j=1}^{b} \gamma_j = 0$$

and

$$(3.1.4) \qquad \sum_{i=1}^{t} \alpha_i^2 = \sum_{j=1}^{b} \gamma_j^2 = 1 \quad .$$

This model is an extension of some known two-way classification models. The most well known of these is the two-way classification model assumed by Ward and Dick (1952). This model is

$$(3.1.5) \qquad y_{ij} = \mu + \tau_i + \beta_j + \delta\tau_i\beta_j + e_{ij} \quad i = 1, 2, \cdots, t; \; j = 1, 2, \cdots, b$$

where it is assumed that $\sum_{i=1}^{t} \tau_i = \sum_{j=1}^{b} \beta_j = 0$ and that $e_{ij} \sim$ i.i.d. $N(0, \sigma^2)$. Ward and Dick gave an iterative procedure to solve the normal equations associated with the model (3.1.5) and also showed that the sum of squares due to the hypothesis that $\delta = 0$ after one iteration is equal to the single degree of freedom sum of squares for non-additivity proposed by Tukey (1949). If one lets

$$\alpha_i = \frac{\tau_i}{\sqrt{\sum_{i=1}^{t} \tau_i^2}} \quad i = 1, =, \cdots, t; \quad \gamma_j = \frac{\beta_j}{\sqrt{\sum_{j=1}^{b} \beta_j^2}} \quad j = 1, 2, \cdots, b,$$

and $\lambda = \delta\left(\sum_{i=1}^{t} \tau_i^2 \sum_{j=1}^{b} \beta_j^2\right)^{\frac{1}{2}}$, it is seen that the model (3.1.5) is special case of the model (3.1.1). A disadvantage of the model (3.1.5) is that it does not in general allow for the possibility that some of the treatment and block combinations may contain non-additivity while others do not. Such is the case when the non-additivity is caused by one or more outliers. Such may also be the case when the non-additivity is caused by some third unknown factor affecting some of the cells. Gollob (1968) remarks that the model (3.1.5) and the following model (3.1.6) possess the limitation that they can not provide useful information unless there are substantial differences between row means or between column means, or between both.

Another model which is a simplification of the model (3.1.1) is Mandel's bundle of straight lines model. The model assumed by Mandel was

(3.1.6) $$y_{ij} = \mu + \tau_i + \beta_j + \delta_i \beta_j + e_{ij}$$

where it was assumed that $\sum_{i=1}^{t} \tau_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{t} \delta_i = 0$ and

$e_{ij} \sim$ i.i.d. $N(0, \sigma^2)$. If one lets

$$\alpha_i = \frac{\delta_i}{\left(\sum_{i=1}^{t} \delta_i^2\right)^{\frac{1}{2}}} \quad i = 1, 2, \cdots, t \,, \quad \gamma_j = \frac{\beta_j}{\left(\sum_{j=1}^{b} \beta_j^2\right)^{\frac{1}{2}}} \quad j = 1, 2, \cdots, b \,, \text{ and}$$

$\lambda = \left(\sum_{i=1}^{t} \delta_i^2\right)^{\frac{1}{2}} \left(\sum_{j=1}^{b} \beta_j^2\right)^{\frac{1}{2}}$, it is seen that Mandel's model is a special case of the model (3.1.1).

The two-way classification model in its most general form is

(3.1.7) $$y_{ij} = \mu_{ij} + e_{ij} \quad \text{where} \quad e_{ij} \sim \text{i.i.d. } N(0, \sigma^2) \,.$$

If one were to assume that there exist μ, τ's, β's, α's, γ's, and λ satisfying the restrictions (3.1.2), (3.1.3) and (3.1.4) such that $\mu_{ij} = \mu + \tau_i + \beta_j + \lambda \alpha_i \gamma_j$, one is interested in determining whether there exists a solution for τ's, β's, α's, γ's, μ and λ in terms of μ_{ij}. This is the context of the following theorem.

Theorem 3.1.1 Let μ_{ij} $i = 1, 2, \cdots, t$; $j = 1, 2, \cdots, b$ be such that

(3.1.8) $$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda \alpha_i \gamma_j \quad i = 1, 2, \cdots, t \,; \; j = 1, 2, \cdots, b$$

where the τ's, β's, α's, and γ's satisfy the restrictions (3.1.2), (3.1.3), and (3.1.4). Then

(3.1.9)

$$\mu = \bar{\mu}_{..} = \frac{1}{bt} \sum_{i=1}^{t} \sum_{j=1}^{b} \mu_{ij} ,$$

$$\tau_i = \bar{\mu}_{i.} - \bar{\mu}_{..} = \frac{1}{b} \sum_{j=1}^{b} \mu_{ij} - \bar{\mu}_{..} \quad i = 1, 2, \cdots, t ,$$

$$\beta_j = \bar{\mu}_{.j} - \bar{\mu}_{..} = \frac{1}{t} \sum_{i=1}^{t} \mu_{ij} - \bar{\mu}_{..} \quad j = 1, 2, \cdots, b ,$$

$$\lambda^2 = Ch_{max} \left[(I_b - \frac{1}{b} J_b) M' (I_t - \frac{1}{t} J_t) M (I_b - \frac{1}{b} J_b) \right]$$

where

$$M = (\mu_{ij}) ,$$

$\underline{\gamma}$ is a characteristic vector of

$(I_b - \frac{1}{b} J_b) M' (I_t - \frac{1}{t} J_t) M (I_b - \frac{1}{b} J_b)$ corresponding to the root λ^2

and $\underline{\alpha}$ is a characteristic vector of

$(I_t - \frac{1}{t} J_t) M (I_b - \frac{1}{b} J_b) M' (I_t - \frac{1}{t} J_t)$ corresponding to the root λ^2 .

Proof: Rewriting (3.1.8) in terms of the matrix M one has

(3.1.10) $$M = \mu \underline{j}_t \underline{j}_b' + \underline{\tau} \underline{j}_b' + \underline{j}_t \underline{\beta}' + \lambda \underline{\alpha} \underline{\gamma}' ,$$

hence $\underline{j}_t' M \underline{j}_b = bt\mu$ which implies

(3.1.11) $$\mu = \bar{\mu}_{..} = \frac{1}{bt} \underline{j}_t' M \underline{j}_b = \frac{1}{bt} \sum_{i=1}^{t} \sum_{j=1}^{b} \mu_{ij} .$$

Multiplying (3.1.10) on left by $\underline{j}_t'$ one obtains $\underline{j}_t' M = t\mu \underline{j}_b' + t\underline{\beta}'$ which implies

(3.1.12)

$$\underline{\beta} = \frac{1}{t} M' \underline{j}_t - \mu \underline{j}_b = \frac{1}{t} M' \underline{j}_t - \bar{\mu}\,.\,. \underline{j}_b = \frac{1}{t}(I_b - \frac{1}{b} J_b) M' \underline{j}_t .$$

Similarly by multiplying (3.1.10) on the right by $\underline{j}_b$ one obtains

$$(3.1.13) \qquad \underline{\tau} = \frac{1}{b} M \underline{j}_b - \bar{\mu}\,.\,. \underline{j}_t = \frac{1}{b}(I_t - \frac{1}{t} J_t) M \underline{j}_b .$$

Now

$$\lambda \underline{\alpha} \underline{\gamma}' = M - \mu \underline{j}_t \underline{j}_b' - \underline{\tau} \underline{j}_b' - \underline{j}_t \underline{\beta}'$$

$$= M - \frac{1}{bt} \underline{j}_t \underline{j}_t' M \underline{j}_b \underline{j}_b' - \frac{1}{b}(I_t - \frac{1}{t} J_t) M \underline{j}_b \underline{j}_b'$$

$$- \frac{1}{t} \underline{j}_t \underline{j}_t' M (I_b - \frac{1}{b} J_b)$$

$$= M - \frac{1}{bt} J_t M J_b - \frac{1}{b}(I_t - \frac{1}{t} J_t) M J_b$$

$$- \frac{1}{t} J_t M (I_b - \frac{1}{b} J_b)$$

$$(3.1.4) \qquad = (I_t - \frac{1}{t} J_t) M (I_b - \frac{1}{b} J_b) = M^* \text{ (say)} .$$

If one multiplies (3.1.14) on the left by $\underline{\alpha}'$ and on the right by $\underline{\gamma}$, one obtains

$$\lambda \underline{\alpha}' \underline{\alpha} \underline{\gamma}' \underline{\gamma} = \underline{\alpha}' (I_t - \frac{1}{t} J_t) M (I_b - \frac{1}{b} J_b) \underline{\gamma}$$

which implies

$$(3.1.15) \qquad \lambda = \underline{\alpha}' M^* \underline{\gamma} = \underline{\alpha}' M \underline{\gamma} .$$

Multiplying (3.1.14) on the left by $\underline{\alpha}'$ gives

$$\lambda \underline{\gamma}' = \underline{\alpha}' M^* \tag{3.1.16}$$

and multiplying (3.1.14) on the right by $\underline{\gamma}$ gives

$$\lambda \underline{\alpha} = M^* \underline{\gamma} \ . \tag{3.1.17}$$

Substituting (3.1.16) in (3.1.17) one obtains

$$\lambda^2 \underline{\alpha} = M^* M^{*\prime} \underline{\alpha} \ , \tag{3.1.18}$$

hence λ^2 is a characteristic root of $M^* M^{*\prime}$ and $\underline{\alpha}$ is the corresponding characteristic vector of $M^* M^{*\prime}$. Similarly it can be shown that $\underline{\gamma}$ is a characteristic vector of $M^{*\prime} M^*$ corresponding to the characteristic root λ^2 of $M^{*\prime} M^*$. It remains to be shown that λ^2 is the largest characteristic root of $M^* M^{*\prime}$. By (3.1.14), $\lambda \underline{\alpha} \underline{\gamma}' = M^*$, thus rank $(M^*) \leq 1$ which implies $M^* M^{*\prime}$ has at most one nonzero characteristic root. $\lambda = 0$ implies $M^* = 0$ which implies rank $(M^*) = 0$ which implies all of the characteristic roots of M^* are zero. In either case λ^2 is the largest characteristic root of $M^* M^{*\prime}$ and the proof is completed.

3.2 Maximum Likelihood Estimates

In this section the maximum likelihood estimates of the parameters in model (3.1.1) will be found. It will be convenient to use a $\hat{}$ over the parameter to indicate the maximum likelihood estimate of the parameter indicated. For example, $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2. Again let $Z = (z_{ij})$ where

$$z_{ij} = y_{ij} - y_{i\cdot} - y_{\cdot j} + y_{\cdot\cdot} \quad i = 1, 2, \cdots, t \ ; \ j = 1, 2, \cdots, b$$

and let $b \leq t$. Let $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_{b-1}$ be the nonzero characteristic roots of $Z'Z$.

Theorem 3.2.1 In the model (3.1.1) the maximum likelihood estimates of the parameters are :

$\hat{\mu} = y_{..}$

$\hat{\tau}_i = y_{i.} - y_{..} \quad i = 1, 2, \cdots, t$

$\hat{\beta}_j = y_{.j} - y_{..} \quad j = 1, 2, \cdots, b$

$\hat{\lambda}^2 = \ell_1$

$\hat{\alpha}$ = a characteristic vector of ZZ' corresponding to the root ℓ_1.

$\hat{\underline{\gamma}}$ = a characteristic vector of $Z'Z$ corresponding to the root ℓ_1.

$\hat{\sigma}^2 = (\ell_2 + \ell_3 + \cdots + \ell_{b-1})/bt = (\mathrm{tr}(Z'Z) - \ell_1)/bt$.

Proof: Let $\underline{\theta}' = [\mu, \tau', \beta', \lambda, \underline{\alpha}', \underline{\gamma}']$. Then the likelihood function of $\underline{\theta}$ is

$$L(\underline{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \mu - \tau_i - \beta_j - \lambda\alpha_i\gamma_j)^2\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \mu - \tau_i - \beta_j)^2 - 2\lambda\sum_{i=1}^{t}\sum_{j=1}^{b}\alpha_i\gamma_j(y_{ij} - \mu - \tau_i - \beta_j) + \lambda^2\sum_{i=1}^{t}\sum_{j=1}^{b}\alpha_i^2\gamma_i^2\right]\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \mu - \tau_i - \beta_j)^2 \right.\right.$$

$$\left.\left. - 2\lambda \sum_{i=1}^{t}\sum_{j=1}^{b} \alpha_i \gamma_j y_{ij} + \lambda^2\right]\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \mu - \tau_i - \beta_j)^2 \right.\right.$$

$$\left.\left. + \lambda^2 - 2\lambda\, \underline{\alpha}'\, Y\, \underline{\gamma}\right]\right\}$$

(3.2.2)
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t}\sum_{j=1}^{b}[(y_{ij} - y_{i.} - y_{.j} + y_{..}) \right.\right.$$

$$- (\mu - y_{..}) - (\tau_i - y_{i.} + y_{..}) - (\beta_j - y_{.j} + y_{..})]^2$$

$$\left.\left. + \lambda^2 - 2\lambda\, \underline{\alpha}'\, Z\, \underline{\gamma}\right]\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2 + bt(\mu - y_{..})^2 \right.\right.$$

$$+ b\sum_{i=1}^{t}(\tau_i - y_{i.} + y_{..})^2 + t\sum_{j=1}^{b}(\beta_j - y_{.j} + y_{..})^2$$

$$\left.\left. + (\lambda - \underline{\alpha}'\, Z\, \underline{\gamma})^2 - (\underline{\alpha}'\, Z\, \underline{\gamma})^2\right]\right\}$$

(3.2.3)
$$\leq \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2 - (\underline{\alpha}'\, Z\, \underline{\gamma})^2\right]\right\}.$$

Now since $e^{u} \geq 1 + u$ for every u one obtains

$$e^{-1} \geq (u + 1)\, e^{-(u + 1)}$$

and thus

$$v e^{-1} \geq v(u + 1)\, e^{-(u + 1)}$$

for all $v \geq 0$ and

$$(3.2.4) \qquad (v e^{-1})^{\frac{n}{2}} \geq \left[v(u+1)\, e^{-(u+1)} \right]^{\frac{n}{2}} \quad \text{for} \quad n = 1, 2, \cdots .$$

In (3.2.3) let

$$v = \frac{bt}{2\pi \left(\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^{2} - (\underline{\alpha}' Z \underline{\gamma})^{2} \right)},$$

$$u = \frac{\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^{2} - (\underline{\alpha}' Z \underline{\gamma})^{2}}{bt\sigma^{2}} - 1,$$

and $n = bt$, then one obtains

$$\left[\frac{bt}{2\pi \left[\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^{2} - \underline{\alpha}' Z \underline{\gamma})^{2} \right]} \right]^{\frac{bt}{2}} e^{-\frac{bt}{2}}$$

$$(3.2.5) \qquad \geq \left(\frac{1}{2\pi\sigma^{2}} \right)^{\frac{bt}{2}} \exp \left\{ -\frac{1}{2\sigma^{2}} \left[\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^{2} - (\underline{\alpha}' Z \underline{\gamma})^{2} \right] \right\}.$$

Thus from (3.2.3)

$$L(\underline{\theta}) \leq \left[\frac{bt}{2\pi \sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 - (\underline{\alpha}' Z \underline{\gamma})^2} \right]^{\frac{bt}{2}} e^{-\frac{bt}{2}} .$$

Now $(\underline{\alpha}' Z \underline{\gamma})^2 \leq \ell_1$ by Theorem (2.2.5), thus

$$(3.2.6) \qquad L(\theta) \leq \left[\frac{bt}{2\pi \sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 - \ell_1} \right]^{\frac{bt}{2}} e^{-\frac{bt}{2}} .$$

Therefore the likelihood function is bounded above by the expression on the right in (3.2.6), a quantity independent of all parameters. Let

$$\sigma^2 = \frac{\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 - \ell_1}{bt} = \hat{\sigma}^2 , \quad \mu = y_{..} = \hat{\mu} ,$$

$$\tau_i = y_{i.} - y_{..} = \hat{\tau}_i \quad i = 1, 2, \cdots, t , \quad \beta_j = y_{.j} - y_{..} = \hat{\beta}_j \quad j = 1, 2, \cdots, b ,$$

$\lambda^2 = \ell_1$, $\underline{\alpha} = \hat{\underline{\alpha}}$ and $\gamma = \hat{\underline{\gamma}}$ where $\hat{\underline{\alpha}}$ and $\hat{\underline{\gamma}}$ are defined in (3.2.1). Substituting the values in the likelihood function $L(\theta)$ in (3.2.2), one obtains

$$L(\hat{\theta}) = \left(\frac{1}{2\pi \hat{\sigma}^2} \right)^{\frac{bt}{2}} \exp \left\{ - \frac{1}{2\hat{\sigma}^2} \left[\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 + \ell_1 - 2\sqrt{\ell_1}\, \hat{\underline{\alpha}}' Z \hat{\underline{\gamma}} \right] \right\}$$

$$= \left(\frac{1}{2\pi \hat{\sigma}^2} \right)^{\frac{bt}{2}} \exp \left\{ - \frac{1}{2\hat{\sigma}^2} \left[\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 + \ell_1 - 2\ell_1 \right] \right\}$$

since from (2.2.5) one has $\ell_1 = (\hat{\underline{\alpha}}' Z \hat{\underline{\gamma}})^2$. Therefore

(3.2.7)
$$L(\hat{\theta}) = \left(\frac{1}{2\pi \hat{\sigma}^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\hat{\sigma}^2}\left(\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2 - \ell_1\right)\right\}$$

$$= \left(\frac{1}{2\pi \hat{\sigma}^2}\right)^{\frac{bt}{2}} e^{-\frac{bt}{2}} .$$

Hence for the values of the parameters given above the likelihood function attains its upper bound (3.2.6) and the theorem is proved.

3.3 Likelihood Ratio Test of Ho: $\lambda = 0$ vs. Ha: $\lambda \neq 0$

In the model (3.1.1) one is interested in testing whether or not there is non-additivity present in the model. This can be done by testing the hypothesis Ho: $\lambda = 0$ vs. Ha: $\lambda \neq 0$. In this section the likelihood ratio test of Ho will be obtained.

Theorem 3.3.1 In the model (3.1.1), the likelihood ratio test of Ho: $\lambda = 0$ vs. Ha: $\lambda \neq 0$ is

$$\Lambda = \left[\frac{\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2 - \ell_1}{\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2}\right]^{\frac{bt}{2}} .$$

Proof:

$$L(\Omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{bt}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \mu - \tau_i - \beta_j - \lambda\alpha_i\gamma_j)^2\right\} .$$

By (3.2.6) and (3.2.7)

$$(3.3.1) \qquad L(\hat{\Omega}) = \left[\frac{bt}{2\pi \left[\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 - \ell_1 \right]} \right]^{\frac{bt}{2}} e^{-\frac{bt}{2}} .$$

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{bt}{2}} \exp \left\{ - \frac{1}{2\sigma^2} \sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - \mu - \tau_i - \beta_j)^2 \right\} .$$

It is easily shown that

$$(3.3.2) \qquad L(\hat{\omega}) = \left(\frac{bt}{2\pi \sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2} \right)^{\frac{bt}{2}} e^{-\frac{bt}{2}} .$$

Thus

$$(3.3.3) \qquad \Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2 - \ell_1}{\sum_{i=1}^{t} \sum_{j=1}^{b} z_{ij}^2} \right]^{\frac{bt}{2}} .$$

One rejects the $Ho: \lambda = 0$ in favor of $Ha: \lambda \neq 0$ if $\Lambda < K$ where K is such that $P_{Ho} \{ \Lambda < K \} = \alpha$ for a given α. $\Lambda < K$ if and only if

$$\frac{\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2 - \ell_1}{\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}} < K^{2/bt}$$

if and only if

$$\frac{\ell_1}{\sum_{i=1}^{t}\sum_{j=1}^{b} z_{ij}^2} > 1 - K^{2/bt} = K^*$$

if and only if

(3.3.4) $$\frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}} > K^*$$

where K^* is such that

(3.3.5) $$P\left\{\frac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}} > K^*\right\} = \alpha$$

Note that this is the same statistic as given in (2.4.1). The distribution of this statistic under Ho : $\lambda = 0$ is given in (2.4.14).

It has previously been shown that Tukey's single degree of freedom sum of square for non-additivity is always smaller than ℓ_1. It might be interesting to study the conditions under which they are approximately equal. Consider

(3.3.6) $$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda \alpha_i \gamma_j$$

where the restrictions (3.1.2), (3.1.3), and (3.1.4) hold. That is, it is assumed that the true cell means are observed and fit the above model exactly. In this case Tukey's sum of squares for non-additivity is

$$(3.3.7)\quad T = \frac{\left[\sum_{i=1}^{t}\sum_{j=1}^{b}(\bar{\mu}_{i.}-\bar{\mu}_{..})(\bar{\mu}_{.j}-\bar{\mu}_{..})(\mu_{ij}-\bar{\mu}_{i.}-\bar{\mu}_{.j}+\bar{\mu}_{..})\right]^2}{\sum_{i=1}^{t}(\bar{\mu}_{i.}-\bar{\mu}_{..})^2 \sum_{j=1}^{b}(\bar{\mu}_{.j}-\bar{\mu}_{..})^2}$$

where $\bar{\mu}_{i.}$, $\bar{\mu}_{.j}$, and $\bar{\mu}_{..}$ are defined in (3.1.9). Then by Theorem 3.1.1,

$$T = \frac{\left[\sum_{i=1}^{t}\sum_{j=1}^{b}\tau_i\beta_j(\lambda\alpha_i\gamma_j)\right]^2}{\sum_{i=1}^{t}\tau_i^2 \sum_{j=1}^{b}\beta_j^2}$$

$$(3.3.8)\quad = \frac{\lambda^2\left(\sum_{i=1}^{t}\tau_i\alpha_i\right)^2\left(\sum_{j=1}^{b}\beta_j\gamma_j\right)^2}{\sum_{i=1}^{t}\tau_i^2 \sum_{j=1}^{b}\beta_j^2}\ .$$

Since

$$\frac{\left(\sum_{i=1}^{t}\tau_i\alpha_i\right)^2}{\sum_{i=1}^{t}\tau_i^2} \le \frac{\sum_{i=1}^{t}\tau_i^2 \sum_{i=1}^{b}\alpha_i^2}{\sum_{i=1}^{t}\tau_i^2} = \sum_{i=1}^{t}\alpha_i^2 = 1$$

and

$$\frac{\left(\sum_{j=1}^{b} \beta_j \gamma_j\right)^2}{\sum_{j=1}^{b} \beta_j^2} \leq \frac{\sum_{j=1}^{b} \beta_j^2 \sum_{j=1}^{b} \gamma_j^2}{\sum_{j=1}^{b} \beta_j^2} = \sum_{j=1}^{b} \gamma_j^2 = 1,$$

one has $T \leq \lambda^2$. It is interesting to note that Tukey's sum of squares for non-additivity will be small if either (1) λ^2 is small, (2) $\frac{(\sum \alpha_i \tau_i)^2}{\sum \tau_i^2}$ is small, or (3) $\frac{(\sum \beta_j \gamma_j)^2}{\sum \beta_j^2}$ is small. These can be interpreted as follows:

(1) λ^2 is small (actually zero) if and only if there is no interaction in the two-way classification design.

(2) $\frac{\left(\sum_{i=1}^{t} \alpha_i \tau_i\right)^2}{\sum_{i=1}^{t} \tau_i^2}$ is small if and only if there is no relationship between the α's and τ's. One might look at this as pseudo correlation. That is, let $\rho^2_{\alpha,\tau} = \frac{\left(\sum_{i=1}^{t} \alpha_i \tau_i\right)^2}{\sum_{i=1}^{t} \tau_i^2} = \frac{\left(\sum_{i=1}^{t} \alpha_i \tau_i\right)^2}{\sum_{i=1}^{t} \tau_i^2 \sum_{i=1}^{t} \alpha_i^2} \leq 1$. Then

$\rho^2_{\alpha,\tau} = 1$ if and only if $\alpha_i = a\tau_i$ for some a and $\sum_{i=1}^{t} \alpha_i^2 = 1$ implies $a = \frac{1}{\left(\sum_{i=1}^{t} \tau_i^2\right)^{\frac{1}{2}}}$. In the case $\rho^2_{\alpha,\tau} = 1$, there is a very definite relationship between the α's and the τ's.

(3) $\dfrac{\left(\sum_{j=1}^{b} \beta_j \gamma_j\right)^2}{\sum_{j=1}^{b} \beta_j^2}$ is small can be interpreted in the same way as (2).

It is also interesting to note that $T = \lambda^2$ if $\dfrac{\left(\sum_{i=1}^{t} \alpha_i \tau_i\right)^2}{\sum_{i=1}^{t} \tau_i^2} = 1$ and

$\dfrac{\left(\sum_{j=1}^{b} \beta_j \gamma_j\right)^2}{\sum_{j=1}^{b} \beta_j^2} = 1$ which are true if and only if

$$\alpha_i = \frac{\tau_i}{\sqrt{\sum_{i=1}^{t} \tau_i^2}} \quad i = 1, 2, \cdots, t \quad \text{and} \quad \gamma_j = \frac{\beta_j}{\sqrt{\sum_{j=1}^{b} \beta_j^2}} \quad j = 1, 2, \cdots, b .$$

In this case the model can be written as

$$\mu_{ij} = \mu + \tau_i + \beta_j + \frac{\lambda}{\left(\sum_{i=1}^{t} \tau_i^2 \sum_{j=1}^{b} \beta_j^2\right)^{\frac{1}{2}}} \tau_i \beta_j$$

$$= \mu + \tau_i + \beta_j + \lambda^* \tau_i \beta_j ; \text{ i.e. model (3.1.5)} .$$

It is the opinion of this writer that if the interaction is a function of the two sets of treatments, then Tukey's test and tests proposed by Milliken and Graybill (1970) are "good" tests for showing significant non-additivity. However, if the interaction is not a function of the two sets of treatment effects, then Tukey's test is not a "good" test for showing significant interaction. In this case the test proposed in (3.3.3) would be preferred.

An example will be given in the next chapter which tends to support the opinion of the writer.

CHAPTER 4

MORE ON THE DISTRIBUTION OF $\ell_1/(\ell_1+\ell_2+\cdots+\ell_{b-1})$

4.1 Introduction

In this chapter the distribution of

$$u_1 = \frac{\ell_1}{\ell_1+\ell_2+\cdots+\ell_{b-1}} \tag{4.1.1}$$

under Ho will be considered for some particular values of b and t. Note that Ho may be either that all two by two table differences are equal to zero as discussed in Chapter 2 or that $\lambda = 0$ in model (3.1.1), as the distribution of u_1, is the same in both cases. The moments of u_1 will be found for these special cases and upper $\alpha \cdot 100\%$ critical points will be found for the distribution of u_1 under Ho.

In section 4.2 the case $b=3$ and $t \geq 3$ will be discussed. In section 4.3 the cases $b=4$, $t=5$ and $b=4$, $t=7$ are discussed. The proofs in this section illustrate the methods that can be used to look at the distribution for any particular values of b and t, but at this time no general results are available. Because of this an approximation to the distribution of $\frac{\ell_1}{\ell_1+\ell_2+\cdots+\ell_{b-1}}$ under Ho will be proposed in section 4.4. The critical points of this approximation will be compared with the exact critical points for those special cases that have been considered in the earlier sections.

4.2 The Special Case $b=3$

Let ℓ_1 and ℓ_2 be the two nonzero characteristic roots of $Z'Z$ where Z is the $t \times 3$, $t \geq 3$, matrix of residuals from a $t \times 3$ two-way classification model (2.1.1) or (3.1.1). Let

$u_1 = \dfrac{\ell_1}{\ell_1 + \ell_2}$ and $u_2 = \ell_1 + \ell_2$. In section 2.4 it was shown that under Ho u_1 and u_2 are independent. The density of u_1 under Ho is given by (2.4.12) as

(4.2.1)

$$g(u_1) = \frac{\pi\,\Gamma(t-1)}{\Gamma(\frac{t-1}{2})\,\Gamma(\frac{t-2}{2})\,\Gamma(1)\,\Gamma(\frac{1}{2})}\; u_1^{\frac{t-4}{2}}\,(1-u_1)^{\frac{t-4}{2}}\,(2u_1-1)$$

for $\frac{1}{2} < u_1 < 1$. It is easily shown that

$$\frac{\pi\,\Gamma(t-1)}{\Gamma(\frac{t-1}{2})\,\Gamma(\frac{t-2}{2})\,\Gamma(1)\,\Gamma(\frac{1}{2})} = (t-2)\,2^{t-3}\;;$$

thus

(4.2.2) $$g(u_1) = (t-2)\,2^{t-3}\,u_1^{\frac{t-4}{2}}\,(1-u_1)^{\frac{t-4}{2}}\,(2u_1-1)$$

for $\frac{1}{2} < u_1 < 1$.

The first moment of u_1 is

(4.2.3)

$$\mathrm{E}\,u_1 = (t-2)\,2^{t-3}\int_{\frac{1}{2}}^{1} u_1^{\frac{t-2}{2}}\,(1-u_1)^{\frac{t-4}{2}}\,(2u_1-1)\,du_1$$

$$= (t-2)\,2^{t-3}\int_{\frac{1}{2}}^{1} u_1^{\frac{t-2}{2}}\,(1-u_1)^{\frac{t-4}{2}}\,(u_1-(1-u_1))\,du_1$$

$$= (t-2)\,2^{t-3}\left[\int_{\frac{1}{2}}^{1} u_1^{\frac{t}{2}}\,(1-u_1)^{\frac{t-4}{2}}\,du_1 - \int_{\frac{1}{2}}^{1} u_1^{\frac{t-2}{2}}\,(1-u_1)^{\frac{t-2}{2}}\,du_1\right].$$

Integrating once by parts in the first integral on the right of (4.2.3) one obtains

(4.2.4)

$$Eu_1 = (t-2)\,2^{t-3}\left[-\frac{2}{t-2}u_1^{\frac{t}{2}}(1-u_1)^{\frac{t-2}{2}}\Big|_{\frac{1}{2}}^{1}\right.$$

$$\left.+\frac{2}{t-2}\int_{\frac{1}{2}}^{1}u_1^{\frac{t-2}{2}}(1-u_1)^{\frac{t-2}{2}}du_1-\int_{\frac{1}{2}}^{1}u_1^{\frac{t-2}{2}}(1-u_1)^{\frac{t-2}{2}}du_1\right]$$

$$=(t-2)\,2^{t-3}\left[\frac{2}{t-2}\cdot\left(\frac{1}{2}\right)^{\frac{t}{2}}\left(\frac{1}{2}\right)^{\frac{t-2}{2}}+\frac{2}{t-2}\int_{\frac{1}{2}}^{1}u_1^{\frac{t-2}{2}}(1-u_1)^{\frac{t-2}{2}}du_1\right]$$

$$=\frac{1}{2}+2^{t-2}\int_{\frac{1}{2}}^{1}u_1^{\frac{t-2}{2}}(1-u_1)^{\frac{t-2}{2}}du_1\,.$$

Since the integrand of the remaining integral on the right of (4.2.4) is symmetric about $\frac{1}{2}$ in the interval $(0, 1)$,

(4.2.5)

$$\int_{\frac{1}{2}}^{1}u_1^{\frac{t-2}{2}}(1-u_1)^{\frac{t-2}{2}}du_1=\tfrac{1}{2}\int_{0}^{1}u_1^{\frac{t-2}{2}}(1-u_1)^{\frac{t-2}{2}}du_1=\tfrac{1}{2}\beta\left(\frac{t}{2},\frac{t}{2}\right).$$

Substituting (4.2.5) in (4.2.4) one sees that

(4.2.6) $$Eu_1=\tfrac{1}{2}+2^{t-3}\beta\left(\frac{t}{2},\frac{t}{2}\right).$$

The variance of u_1 will now be obtained.

(4.2.6)

$$E[u_1(1-u_1)] = (t-2)\,2^{t-3}\int_{\frac{1}{2}}^{1} u_1^{\frac{t+2-4}{2}}(1-u_1)^{\frac{t+2-4}{2}}(2u_1-1)\,du_1$$

$$= \frac{(t-2)\,2^{t-3}}{t\;2^{t-1}}\int_{\frac{1}{2}}^{1}(t+2-2)\,2^{t+2-3}\,u_1^{\frac{t+2-4}{2}}(1-u_1)^{\frac{t+2-4}{2}}(2u_1-1)\,du_1$$

$$= \frac{(t-2)\,2^{t-3}}{t\;2^{t-1}}\cdot 1 = \frac{t-2}{4t}\quad .$$

Thus (4.2.6) implies that

$$Eu_1 - Eu_1^2 = \frac{t-2}{4t}$$

which implies

$$Eu_1^2 = \frac{1}{2} + 2^{t-3}\beta\left(\frac{t}{2},\frac{t}{2}\right) - \frac{t-2}{4t}$$

which implies

(4.2.7)

$$\text{var}(u_1) = Eu_1^2 - (Eu_1)^2$$

$$= \frac{1}{2} + 2^{t-3}\beta\left(\frac{t}{2},\frac{t}{2}\right) - \frac{t-2}{4t} - \left(\frac{1}{2} + 2^{t-3}\beta\left(\frac{t}{2},\frac{t}{2}\right)\right)^2$$

$$= \left[\frac{1}{2} + 2^{t-3}\beta\left(\frac{t}{2},\frac{t}{2}\right)\right]\left[1 - \frac{1}{2} - 2^{t-3}\beta\left(\frac{t}{2},\frac{t}{2}\right)\right] - \frac{t-2}{4t}$$

$$= \frac{1}{4} - 2^{2t-6}\beta^2\left(\frac{t}{2},\frac{t}{2}\right) - \frac{t-2}{4t}$$

$$= \frac{1}{2t} - 2^{2t-6}\beta^2\left(\frac{t}{2},\frac{t}{2}\right)\quad .$$

Theorem 4.2.1 In the model (2.1.1) with $b = 3$ and with $\gamma_{ij} = 0$ for $i = 1, 2, \cdots, t$; $j = 1, 2, 3$, the following are true:

(4.2.8) $$E\,\ell_1 = [(t-1) + 2^{t-2}(t-1)\beta(\tfrac{t}{2},\tfrac{t}{2})]\,\sigma^2$$

(4.2.9) $$E\,\ell_2 = [(t-1) - 2^{t-2}(t-1)\beta(\tfrac{t}{2},\tfrac{t}{2})]\,\sigma^2$$

(4.2.10) $$\mathrm{var}(\ell_1) = (t-1)[3 + 2^{t-1}\beta(\tfrac{t}{2},\tfrac{t}{2}) - 2^{2t-4}(t-1)\beta^2(\tfrac{t}{2},\tfrac{t}{2})]\,\sigma^4$$

(2.2.11) $$\mathrm{var}(\ell_2) = (t-1)[3 - 2^{t-1}\beta(\tfrac{t}{2},\tfrac{t}{2}) - 2^{2t-4}(t-1)\beta^2(\tfrac{t}{2},\tfrac{t}{2})]\,\sigma^4$$

(4.2.12) $$\mathrm{cov}(\ell_1, \ell_2) = (t-1)[-1 + 2^{2t-4}(t-1)\beta^2(\tfrac{t}{2},\tfrac{t}{2})]\,\sigma^4$$

Proof:

$$E\,\ell_1 = E\left[\frac{\ell_1}{\ell_1+\ell_2}\cdot(\ell_1+\ell_2)\right]$$

$$= E[u_1u_2]$$

$$= Eu_1\ Eu_2$$

since u_1 and u_2 are independent. Note that $u_2 = \ell_1 + \ell_2$

$= \sum_{i=1}^{t}\sum_{j=1}^{3}(y_{ij} - y_{i.} - y_{.j} + y_{..})^2$ and that $\frac{u_2}{\sigma^2} \sim \chi^2(2(t-1))$ under Ho, thus

(4.2.12) $$Eu_2 = 2(t-1)\,\sigma^2$$

and

(4.2.14) $$\mathrm{Var}(u_2) = 4(t-1)\,\sigma^4$$

so that

(4.2.15) $$E u_2^2 = [4(t-1) + 4(t-1)^2]\,\sigma^4$$

$$= 4t(t-1)\sigma^4 .$$

Therefore

$$E \ell_1 = E u_1 \, E u_2$$

$$= \left[\frac{1}{2} + 2^{t-3}\beta(\tfrac{t}{2}, \tfrac{t}{2})\right] \cdot 2(t-1)\sigma^2$$

and (4.2.8) follows.

$$E \ell_2 = E(\ell_1 + \ell_2 - \ell_1)$$

$$= E(\ell_1 + \ell_2) - E\ell_1$$

$$= 2(t-1)\sigma^2 - \left[(t-1) + 2^{t-2}\beta(\tfrac{t}{2}, \tfrac{t}{2})\right]\sigma^2$$

$$= \left[(t-1) - 2^{t-2}(t-1)\beta(\tfrac{t}{2}, \tfrac{t}{2})\right]\sigma^2$$

which proves (4.2.9) .

$$E \ell_1^2 = E\left[\left(\frac{\ell_1}{\ell_1 + \ell_2}\right)^2 (\ell_1 + \ell_2)^2\right]$$

$$= E\left[u_1^2 \, u_2^2\right]$$

$$= E u_1^2 \, E u_2^2$$

$$= \left[\frac{1}{2} + 2^{t-3}\beta(\tfrac{t}{2}, \tfrac{t}{2}) - \frac{t-2}{4t}\right] 4t(t-1)\sigma^4$$

$$= \left[\frac{t+2}{4t} + 2^{t-3}\beta(\tfrac{t}{2}, \tfrac{t}{2})\right] 4t(t-1)\sigma^4$$

$$= \left[t + 2 \; + 2^{t-1} t\beta(\tfrac{t}{2}, \tfrac{t}{2})\right] (t-1)\sigma^4$$

thus

$$\text{var}(\ell_1) = E\ell_1^2 - (E\ell_1)^2$$

$$= \Big\{(t+2)(t-1) + t(t-1)2^{t-1}\beta(\tfrac{t}{2}, \tfrac{t}{2})$$

$$- \left[t - 1 + 2^{t-2}(t-1)\beta(\tfrac{t}{2}, \tfrac{t}{2})\right]^2 \Big\} \sigma^4$$

$$= \Big[t^2 + t - 2 + t(t-1)2^{t-1}\beta(\tfrac{t}{2}, \tfrac{t}{2}) - t^2 + 2t - 1$$

$$- 2^{t-1}(t-1)^2\beta(\tfrac{t}{2}, \tfrac{t}{2}) - 2^{2t-4}(t-1)^2\beta^2(\tfrac{t}{2}, \tfrac{t}{2})\Big] \sigma^4$$

$$= \left[3t - 3 + 2^{t-1}(t-1)\beta(\tfrac{t}{2}, \tfrac{t}{2})(t - t + 1) - 2^{2t-4}(t-1)^2\beta(\tfrac{t}{2}, \tfrac{t}{2})\right] \sigma^4$$

$$= (t-1)\left[3 + 2^{t-1}\beta(\tfrac{t}{2}, \tfrac{t}{2}) - 2^{2t-4}(t-1)\beta^2(\tfrac{t}{2}, \tfrac{t}{2})\right] \sigma^4$$

which proves (4.2.10) .

$$E(\ell_1\ell_2) = E\left[\frac{\ell_1}{\ell_1+\ell_2} \cdot \frac{\ell_2}{\ell_1+\ell_2} \cdot (\ell_1+\ell_2)^2\right]$$

$$= E\left[u_1(1-u_1)\, u_2^2\right]$$

$$= E\left[u_1(1-u_1)\right] E u_2^2$$

$$= \frac{t-2}{4t} \cdot 4t(t-1)\sigma^4$$

$$= (t-1)(t-2)\sigma^4 .$$

Therefore

$$\operatorname{cov}(\ell_1, \ell_2) = \sigma^4\left[(t-1)(t-2) - (t-1)(1+2^{t-2}\beta(\tfrac{t}{2},\tfrac{t}{2}))(t-1)(1-2^{t-2}\beta(\tfrac{t}{2},\tfrac{t}{2}))\right]$$

$$= \sigma^4\left[(t-1)(t-2) - (t-1)^2\,(1-2^{2t-4}\beta^2(\tfrac{t}{2},\tfrac{t}{2}))\right]$$

$$= \sigma^4\left[-(t-1) + 2^{2t-4}(t-1)^2\beta^2(\tfrac{t}{2},\tfrac{t}{2})\right]$$

which proves (4.2.12)

$$\operatorname{var}(\ell_2) = \operatorname{var}(\ell_1 + \ell_2 - \ell_1)$$

$$= \operatorname{var}(\ell_1 + \ell_2) - 2\operatorname{cov}(\ell_1 + \ell_2, \ell_1) + \operatorname{var}(\ell_1)$$

$$= \operatorname{var}(u_2) - 2\operatorname{cov}(\ell_1, \ell_1) - 2\operatorname{cov}(\ell_2, \ell_1) + \operatorname{var}\ell_1$$

$$= \operatorname{var}(u_2) - 2\operatorname{cov}(\ell_1, \ell_2) - \operatorname{var}(\ell_1)$$

$$= \sigma^4\left[4(t-1) + 2(t-1) - 2^{2t-3}(t-1)^2\beta^2(\tfrac{t}{2},\tfrac{t}{2})\right.$$

$$\left. - 3(t-1) - 2(t-1)\,2^{t-1}\beta(\tfrac{t}{2},\tfrac{t}{2}) + 2^{2t-4}(t-1)^2\beta^2(\tfrac{t}{2},\tfrac{t}{2})\right]$$

$$= \sigma^4\left[3(t-1) - 2(t-1)\,2^{t-1}\beta(\tfrac{t}{2},\tfrac{t}{2}) - 2^{2t-4}(t-1)^2\beta^2(\tfrac{t}{2},\tfrac{t}{2})\right]$$

which proves (4.2.11).

Now the upper $\alpha \cdot 100\%$ critical points for the distribution of u_1 under Ho will be found. Note that

$$P_{Ho}\{u_1 > x\} = \alpha$$

if and only if

$$P_{Ho}\{u_1 - u_1 x > x - u_1 x\} = \alpha$$

if and only if

$$P_{Ho}\{u_1(1-x) > x\ (1-u_1)\} = \alpha$$

if and only if

$$P_{Ho}\left\{\frac{u_1}{1-u_1} > \frac{x}{1-x}\right\} = \alpha$$

since $0 < u_1 < 1$ and $0 < x < 1$. From (4.2.2)

$$g(u_1) = (t-2)\,2^{t-3}\,u_1^{\frac{t-4}{2}}\,(1-u_1)^{\frac{t-4}{2}}\,(2u_1 - 1) \quad \text{for} \quad \tfrac{1}{2} < u_1 < 1 .$$

Let $w = \dfrac{u_1}{1-u_1}$ which implies $u_1 = \dfrac{w}{1+w}$, thus the jacobian of the transformation is $\dfrac{1}{(1+w)^2}$. Therefore the density of w is

$$(4.2.16)\quad h(w) = (t-2)\,2^{t-3}\left(\frac{w}{1+w}\right)^{\frac{t-4}{2}}\left(\frac{1}{1+w}\right)^{\frac{t-4}{2}}\left(\frac{w-1}{1+w}\right)\frac{1}{(1+w)^2}$$

for $1 < w < \infty$ which implies

$$h(w) = (t-2)2^{t-3}\,\frac{(w-1)w^{\frac{t-4}{2}}}{(1+w)^{t-1}} \quad \text{for} \quad 1 < w < \infty .$$

Now

$$(4.2.17)\quad \alpha = P_{Ho}\{u_1 > x\} = P_{Ho}\left\{w > \frac{x}{1-x}\right\}$$

$$= \int_{x/1-x}^{\infty} (t-2)\,2^{t-3}\,\frac{(w-1)w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw .$$

(4.2.18)
$$\int_y^\infty \frac{(w-1)\,w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw = \int_y^\infty \frac{w^{\frac{t-2}{2}}}{(1+w)^{t-1}}\,dw - \int_y^\infty \frac{w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw \ .$$

Integrating the first integral on the right of (4.2.18) by parts one obtains

$$\int_y^\infty \frac{(w-1)\,w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw$$

$$= \frac{-w^{\frac{t-2}{2}}}{(t-2)(1+w)^{t-2}}\Bigg|_y^\infty + \int_y^\infty \frac{w^{\frac{t-4}{2}}}{2(1+w)^{t-2}}\,dw - \int_y^\infty \frac{w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw$$

$$= \frac{y^{\frac{t-2}{2}}}{(t-2)(1+y)^{t-2}} + \frac{1}{2}\int_y^\infty \frac{(w-1)\,w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw \ .$$

Thus

(4.2.19)
$$\int_y^\infty \frac{(w-1)\,w^{\frac{t-4}{2}}}{(1+w)^{t-1}}\,dw = \frac{2\,y^{\frac{t-2}{2}}}{(t-2)(1+y)^{t-2}} = \frac{2}{t-2}\left(\frac{y}{(1+y)^2}\right)^{\frac{t-2}{2}} \ .$$

Replacing y by $\frac{x}{1-x}$ in (4.2.19) and substituting this in (4.2.17) one obtains

(4.2.20)
$$\alpha = (t-2)\,2^{t-3}\cdot\frac{2}{t-2}\left[\frac{\frac{x}{1-x}}{\left(1+\frac{x}{1-x}\right)^2}\right]^{\frac{t-2}{2}}$$

$$= 2^{t-2}\,[x(1-x)]^{\frac{t-2}{2}}$$

$$= [4x(1-x)]^{\frac{t-2}{2}} .$$

Therefore

$$P_{Ho}\{u_1 > x\} = \alpha .$$

is equivalent to

$$[4x(1-x)]^{\frac{t-2}{2}} = \alpha$$

which is equivalent to

$$4x^2 - 4x + \alpha^{\frac{2}{t-2}} = 0 .$$

Thus

(4.2.21)
$$x = \frac{4 \pm \sqrt{16 - 16\alpha^{\frac{2}{t-2}}}}{8}$$

$$= \frac{1 \pm \sqrt{1 - \alpha^{\frac{2}{t-2}}}}{2} .$$

The root $\frac{1 - \sqrt{1 - \alpha^{\frac{2}{t-2}}}}{2}$ in (4.2.21) is less than $\frac{1}{2}$ and hence is an extraneous root since

$$P_{Ho}\{u_1 \geq \frac{1 - \sqrt{1 - \alpha^{\frac{2}{t-2}}}}{2}\} \geq P_{Ho}\{u_1 \geq \tfrac{1}{2}\} = 1 \neq \alpha .$$

Therefore the upper $\alpha \cdot 100\%$ critical point for the distribution of u_1 under Ho is

(4.2.22)
$$x = \frac{1 + \sqrt{1 - \alpha^{\frac{2}{t-2}}}}{2} .$$

4.3 The Special Case $b = 4$

Let ℓ_1, ℓ_2, ℓ_3 be the three nonzero characteristic roots of $Z'Z$ where Z is the $t \times 4 \ (t \geq 4)$ matrix of residuals from a $t \times 4$ two-way classification model (2.1.1) or (3.1.1). Let

$$u_1 = \frac{\ell_1}{\ell_1 + \ell_2 + \ell_3}, \quad u_2 = \frac{\ell_2}{\ell_1 + \ell_2 + \ell_3}, \text{ and } \quad u_3 = \ell_1 + \ell_2 + \ell_3 .$$

In section 2.4 it was shown that u_1 and u_2 are independent of u_3. Instead of working with the density of u_1, it will be easier to work with the density of

$$v_1 = \frac{3u_1 - 1}{2} .$$

The density of v_1 will now be obtained.

From (2.4.10) the joint density of u_1 and u_2 is

(4.3.1)
$$g(u_1, u_2) = K[u_1 u_2 (1 - u_1 - u_2)]^{\frac{t-5}{2}} (u_1 - u_2)(2u_1 + u_2 - 1)(u_1 + 2u_2 - 1)$$

for $\frac{1}{3} < u_1 < 1$ and $\frac{1 - u_1}{2} < u_2 < \min(u_1, 1 - u_1)$ where

$$K = \frac{\pi^{\frac{3}{2}} \Gamma\left(\frac{3t-3}{2}\right)}{\Gamma(\frac{t-1}{2})\Gamma(\frac{t-2}{2})\Gamma(\frac{t-3}{2})\Gamma(\frac{3}{2})\Gamma(\frac{2}{2})\Gamma(\frac{1}{2})}$$

(4.3.2)
$$= \frac{2\sqrt{\pi}\ \Gamma\left(\frac{3t-3}{2}\right)}{\Gamma(\frac{t-1}{2})\Gamma(\frac{t-2}{2})\Gamma(\frac{t-3}{2})} .$$

Let $v_1 = \frac{3u_1 - 1}{2}$ and $v_2 = u_2 - \frac{1 - u_1}{2} = \frac{u_1 + 2u_2 - 1}{2}$. This implies

$u_1 = \frac{2v_1 + 1}{3}$ and $u_2 = \frac{1 - v_1 + 3v_2}{3}$. Thus the jabobian of the

of the transformation is

$$J = \begin{vmatrix} \frac{2}{3} & 0 \\ -\frac{1}{3} & 1 \end{vmatrix} = \frac{2}{3} \quad .$$

The range is transformed as follows :

$$\frac{1}{3} < u_1 < 1 \quad \text{and} \quad \frac{1 - u_1}{2} < u_2 < \min(u_1, u_2)$$

implies

$$\frac{1}{3} < \frac{2v_1 + 1}{3} < 1 \quad \text{and} \quad \frac{1}{2} - \frac{2v_1 + 1}{6} < \frac{1 - v_1 + 3v_2}{3}$$

$$< \min\left(\frac{2v_1 + 1}{3} , \frac{2 - 2v_1}{3}\right)$$

which implies

$$0 < v_1 < 1 \quad \text{and} \quad \frac{1 - v_1}{3} < v_2 + \frac{1 - v_1}{3} < \min\left(\frac{2v_1 + 1}{3} , \frac{2(1 - v_1)}{3}\right)$$

which implies

$$(4.3.3) \qquad 0 < v_1 < 1 \quad \text{and} \quad 0 < v_2 < \min\left(v_1, \frac{1 - v_1}{3}\right) .$$

Therefore the joint density of v_1 and v_2 is

$$h(v_1, v_2) = K\left[\frac{2v_1 + 1}{3} \cdot \frac{1 - v_1 + 3v_2}{3} \cdot \frac{1 - v_1 - 3v_2}{3}\right]^{\frac{t-5}{2}}$$

$$\cdot (v_1 - v_2)(v_1 + v_2)(2v_2) \cdot \frac{2}{3}$$

$$= \frac{4K}{3^{\frac{3t-13}{2}}} (2v_1 + 1)^{\frac{t-5}{2}} v_2 [(1 - v_1)^2 - 9v_2^2]^{\frac{t-5}{2}} (v_1^2 - v_2^2)$$

for $0 < v_1 < 1$ and $0 < v_2 < \min\left(v_1, \frac{1 - v_1}{3}\right)$ where K is defined in (4.3.2).

To get the marginal density of v_1, one must integrate over v_2. Consider the integral

$$I = \int_0^M v_2 \left[(1 - v_1)^2 - 9v_2^2\right]^{\frac{t-5}{2}} (v_1^2 - v_2^2)\, dv_2 .$$

Integrate this once by parts by letting

$$x = v_1^2 - v_2^2 \quad \text{and} \quad dy = v_2 \left[(1 - v_1)^2 - 9v_2^2\right]^{\frac{t-5}{2}} dv_2 .$$

Thus

$$dx = -2v_2 dv_2 \quad \text{and} \quad y = \frac{-2}{18(t-3)} \left[(1 - v_1)^2 - 9v_2^2\right]^{\frac{t-3}{2}}$$

and therefore

$$I = \frac{-(v_1^2 - v_2^2)}{9(t-3)} \left[(1 - v_1)^2 - 9v_2^2\right]^{\frac{t-3}{2}} \Bigg|_0^M$$

$$- \int_0^M \frac{2v_2}{9(t-3)} \left[(1 - v_1)^2 - 9v_2^2\right]^{\frac{t-3}{2}} dv_2$$

$$= \frac{v_1^2 (1 - v_1)^{t-3} - (v_1^2 - M^2)\left[(1 - v_1)^2 - 9M^2\right]^{\frac{t-1}{2}}}{9(t-3)}$$

$$- \frac{2}{9(t-3)} \left[- \frac{2}{18(t-1)} \left[(1 - v_1)^2 - 9v_2^2\right]^{\frac{t-1}{2}} \Bigg|_0^M \right.$$

$$(4.3.5) \quad = \frac{v_1^2(1-v_1)^{t-3} - (v_1 - M^2)\left[(1-v_1)^2 - 9M^2\right]^{\frac{t-3}{2}}}{9(t-3)}$$

$$+ \frac{2\left[(1-v_1)^2 - 9M^2\right]^{\frac{t-1}{2}} - 2(1-v_1)^{t-1}}{81\,(t-1)(t-3)} .$$

Let $M = \min\left(v_1, \frac{1-v_1}{3}\right)$, then

$M = v_1$ if and only if $0 < v_1 \leq \frac{1}{4}$ and

$M = \frac{1-v_1}{3}$ if and only if $\frac{1}{4} \leq v_1 < 1$.

When $M = v_1$, (4.3.5) becomes

$$(4.3.6) \quad I = \frac{v_1^2(1-v_1)^{t-3}}{9(t-3)} + \frac{2(1-2v_1-8v_1^2)^{\frac{t-1}{2}} - 2(1-v_1)^{t-1}}{81(t-1)(t-3)}$$

$$= \frac{(1-v_1)^{t-3}}{9(t-3)}\left[v_1^2 - \frac{2(1-v_1^2)^2}{9(t-1)}\right] + \frac{2(1-2v_1-8v_1^2)^{\frac{t-1}{2}}}{81(t-1)(t-3)}$$

and when $M = \frac{1-v_1}{3}$, (4.3.5) becomes

$$(4.3.7) \quad I = \frac{v_1^2(1-v_1)^{t-3}}{9(t-3)} - \frac{2(1-v_1)^{t-1}}{81(t-1)(t-3)}$$

$$= \frac{(1 - v_1)^{t-3}}{9(t-3)} \left[v_1^2 - \frac{2(1 - v_1)^2}{9(t-1)} \right] .$$

Therefore from (4.3.4) the density of v_1 is

$$h_1(v_1) = \frac{4K}{3^{\frac{3t-9}{2}}} \frac{(2v_1 + 1)^{\frac{t-5}{2}}}{t-3} \left[(1 - v_1)^{t-3} \left[v_1^2 - \frac{2(1 - v_1)^2}{9(t-1)} \right] + \frac{2(1 - 2v_1 - 8v_1^2)^{\frac{t-1}{2}}}{9(t-1)} \right] \quad \text{for } 0 < v_1 \leq \tfrac{1}{4}$$

$$= \frac{4K}{3^{\frac{3t-9}{2}}} \frac{(2v_1 + 1)^{\frac{t-5}{2}}}{t-3} \left[(1 - v_1)^{t-3} \left[v_1^2 - \frac{2(1 - v_1)^2}{9(t-1)} \right] \right]$$

for $\frac{1}{4} < v_1 < 1$ which implies

(4.3.8)

$$h_1(v_1) = C(2v_1 + 1)^{\frac{t-5}{2}} \left[(1 - v_1)^{t-3} (9tv_1^2 - 11v_1^2 + 4v_1 - 2) + 2(1 - 2v_1 - 8v_1^2)^{\frac{t-1}{2}} \right] \quad \text{for } 0 < v_1 \leq \tfrac{1}{4}$$

$$= C(2v_1 + 1)^{\frac{t-5}{2}} (1 - v_1)^{t-3} (9tv_1^2 - 11v_1^2 + 4v_1 - 2)$$

for $\frac{1}{4} < v_1 < 1$

where

(4.3.9)

$$C = \frac{4K}{3^{\frac{3t-9}{2}}\,(t-3)(t-1)\cdot 9}$$

$$= \frac{8\sqrt{\pi}\;\Gamma\left(\frac{3t-3}{2}\right)}{3^{\frac{3t-5}{2}}\;\Gamma\left(\frac{t-1}{2}\right)\Gamma\left(\frac{t-2}{2}\right)\Gamma\left(\frac{t-3}{2}\right)(t-1)(t-3)}$$

$$= \frac{2\sqrt{\pi}\;\Gamma\left(\frac{3t-3}{2}\right)}{3^{\frac{3t-5}{2}}\;\Gamma\left(\frac{t+1}{2}\right)\Gamma\left(\frac{t-2}{2}\right)\Gamma\left(\frac{t-1}{2}\right)} \quad .$$

Moments of v_1 and ℓ_1 and critical points of v_1 will be evaluated for the special cases $t = 5$ and $t = 7$.

First consider the density (4.3.8) when $t = 5$. From 4.3.9

$$C = \frac{2\sqrt{\pi}\;\Gamma(6)}{3^5\,\Gamma(3)\,\Gamma(\frac{3}{2})\,\Gamma(2)} = \frac{80}{81} \quad .$$

Hence

$$h_1(v_1) = \frac{80}{81}\left[(1-v_1)^2(34v_1^2 + 4v_1 - 2) + 2(1 - 2v_1 - 8v_1^2)^2\right]$$

$$\text{for} \quad 0 < v_1 \leq \tfrac{1}{4}$$

$$= \frac{80}{81}\left[(1-v_1)^2(34v_1^2 + 4v_1 - 2)\right] \quad \text{for} \quad \tfrac{1}{4} < v_1 < 1$$

which implies

$$h_1(v_1) = \frac{160}{81}\left[(1-v_1)^2(17v_1^2+2v_1-1) + (1-2v_1-8v_1^2)^2\right]$$

$$\text{for} \quad 0 < v_1 \leq \tfrac{1}{4}$$

$$= \frac{160}{81}(1-v_1)^2(17v_1^2+2v_1-1) \quad \text{for } \tfrac{1}{4} < v_1 < 1 \ .$$

The first two moments of v_1 will now be obtained.

$$Ev_1 = \frac{160}{81}\int_0^1 v_1(1-v_1)^2(17v_1+2v_1-1)\,dv_1$$

$$+ \frac{160}{81}\int_0^{\frac{1}{4}} v_1(1-2v_1-8v_1^2)^2\,dv_1$$

$$= \frac{160}{81}\int_0^1 (17v_1^5 - 32v_1^4 + 12v_1^3 + 4v_1^2 - v_1)\,dv_1$$

$$+ \frac{160}{81}\int_0^{\frac{1}{4}} (64v_1^5 + 32v_1^4 - 12v_1^3 - 4v_1^2 + v_1)\,dv_1$$

$$= \frac{160}{81}\left[\frac{17}{6} - \frac{32}{5} + \frac{12}{4} + \frac{4}{3} - \frac{1}{2} + \frac{64}{6\cdot 2^{12}} + \frac{32}{5\cdot 2^{10}}\right.$$

$$\left. - \frac{12}{4\cdot 2^8} - \frac{4}{3\cdot 2^6} + \frac{1}{2\cdot 2^4}\right]$$

$$(4.3.11) \qquad = \frac{13}{24} \ .$$

$$E v_1^2 = \frac{160}{81} \int_0^1 (17v_1^6 - 32v_1^5 + 12v_1^4 + 4v_1^3 - v_1^2)\, dv_1$$

$$+ \frac{160}{81} \int_0^{\frac{1}{4}} (64v_1^6 + 32v_1^5 - 12v_1^4 - 4v_1^3 + v_1^2)\, dv_1$$

$$= \frac{160}{81} \left[\frac{17}{7} - \frac{32}{6} + \frac{12}{5} + \frac{4}{4} - \frac{1}{3} + \frac{64}{7 \cdot 2^{14}} + \frac{32}{6 \cdot 2^{12}} \right.$$

$$\left. - \frac{12}{5 \cdot 2^{10}} - \frac{4}{4 \cdot 2^8} + \frac{1}{3 \cdot 2^6} \right]$$

(4. 3. 12) $$= \frac{9}{28} .$$

$$\text{var}(v_1) = E v_1^2 - (E v_1)^2$$

$$= \frac{9}{28} - \left(\frac{13}{24}\right)^2 = \frac{113}{4032} = .028026 .$$

Since $u_1 = \frac{2v_1 + 1}{3}$ one obtains

(4.3.13) $$E u_1 = \frac{25}{36}$$

and

(4.3.14) $$E u_1^2 = E\left(\frac{4v_1^2 + 4v_1 + 1}{9}\right) = \frac{187}{378} .$$

Now

$$\frac{u_3}{\sigma^2} = \frac{\ell_1 + \ell_2 + \ell_3}{\sigma^2} \sim x^2(12)$$

so that

(4.3.15) $\quad Eu_3 = 12\sigma^2$, $\operatorname{var}(u_3) = 24\sigma^4$, and $Eu_3^2 = 168\sigma^4$.

Thus

$$E\,\ell_1 = E\left[\frac{\ell_1}{\ell_1+\ell_2+\ell_3}\cdot(\ell_1+\ell_2+\ell_3)\right]$$

$$= E[u_1 u_3]$$

$$= Eu_1\,Eu_3 = \frac{25}{36}\cdot 12\sigma^2$$

from (4.3.12) and (4.3.15). Hence

(4.3.16) $\quad E\,\ell_1 = \frac{25}{3}\sigma^2$.

Also

$$E\,\ell_1^2 = E\left[\frac{\ell_1^2}{(\ell_1+\ell_2+\ell_3)^2}\cdot(\ell_1+\ell_2+\ell_3)^2\right]$$

$$= E[u_1^2 u_3^2]$$

$$= Eu_1^2\,Eu_3^2 = \frac{187}{378}\cdot 168\sigma^4$$

from (4.3.14) and (4.3.15). So that

$$E\,\ell_1^2 = \frac{748}{9}\sigma^4 .$$

Thus

(4.3.17) $\quad \operatorname{var}(\ell_1) = \left[\frac{748}{9} - \left(\frac{25}{3}\right)^2\right]\sigma^4 = \frac{41}{3}\sigma^4$

and

(4.3.18) $$S.D.(\ell_1) = 3.682\sigma^2 \quad .$$

Now the upper $\alpha \cdot 100\%$ critical point for the distribution of v_1 will be obtained.

$$P_{Ho}\{v_1 < x\} = 1 - \alpha$$

is equivalent to

$$\int_0^x h_1(v_1)\, dv_1 = 1 - \alpha \quad .$$

Hence

$$1 - \alpha = \frac{160}{81}\left[\int_0^{\frac{1}{4}} (64v_1^4 + 32v_1^3 - 12v_1^2 - 4v_1 + 1)\, dv_1 \right.$$

$$\left. + \int_0^x (17v_1^4 - 32v_1^3 + 12v_1^2 + 4v_1 - 1)\, dv_1\right]$$

unless the first integral in the brackets above is larger than $\frac{81}{160}(1-\alpha)$. Thus

$$1 - \alpha = \frac{160}{81}\left[\frac{64}{5 \cdot 2^{10}} + \frac{32}{4 \cdot 2^8} - \frac{12}{3 \cdot 2^6} - \frac{4}{2 \cdot 2^4} + \frac{17x^5}{5}\right.$$

$$\left. - \frac{32x^4}{4} + \frac{12x^3}{3} + \frac{4x^2}{2} - x\right]$$

$$= \frac{160}{81}\left[\frac{1}{80} + \frac{1}{32} - \frac{1}{16} - \frac{1}{8} + \frac{1}{4} + \frac{17x^5}{5} - 8x^4 + 4x^3\right.$$

$$\left. + 2x^2 - x\right]$$

$$= \frac{17}{81} + \frac{32}{81}\,(17x^5 - 40x^4 + 20x^3 + 10x^2 - 5x)$$

which implies

$$17x^5 - 40x^4 + 20x^3 + 10x^2 - 5x - \frac{81(1-\alpha) - 17}{32} = 0 \quad \text{if}$$

$$1 - \alpha > \frac{17}{81} .$$

Solving the above equation for x when $\alpha = .10, .05, .01$ one obtains

(4.3.19) $$x_{.10} = .7689,\ x_{.05} = .8222,\ \text{and}\ x_{.01} = .8963$$

as the upper critical points when $b = 4$ and $t = 5$.

Now consider the density (4.3.8) when $t = 7$. From (4.3.9)

$$C = \frac{2\sqrt{\pi}\ \Gamma(9)}{3^8\,\Gamma(4)\,\Gamma(\frac{5}{2})\,\Gamma(3)} = \frac{2^8 \cdot 7 \cdot 5}{3^8}$$

and from (4.3.8)

$$h_1(v_1) = \frac{2^8 \cdot 7 \cdot 5}{3^8}(2v_1 + 1)\left[(1 - v_1)^4(52v_1^2 + 4v_1 - 2)\right.$$

$$\left. + 2(1 - 2v_1 - 8v_1^2)^3\right] \qquad \text{for } 0 < v_1 \le \tfrac{1}{4}$$

$$= \frac{2^8 \cdot 7 \cdot 5}{3^8}(2v_1 + 1)(1 - v_1)^4(52v_1^2 + 4v_1 - 2)$$

$$\text{for } \tfrac{1}{4} < v_1 < 1 .$$

Thus

$$h_1(v_1) = \frac{2^9 \cdot 7 \cdot 5}{3^8} [\, (52v_1^7 - 178v_1^6 + 192v_1^5 - 29v_1^4 - 64v_1^3$$

$$+ 24v_1^2 + 4v_1 - 1) \; + \; (1 - 4v_1 - 24v_1^2 + 64v_1^3$$

$$+ 272v_1^4 - 192v_1^5 - 1280v_1^6 - 1024v_1^7) \,]$$

$$\text{for} \quad 0 < v_1 \leq \tfrac{1}{4}$$

$$= \frac{2^9 \cdot 7 \cdot 5}{3^8} (52v_1^7 - 178v_1^6 + 192v_1^5 - 29v_1^4 - 64v_1^3$$

$$+ 24v_1^2 + 4v_1 - 1) \quad \text{for} \quad \tfrac{1}{4} < v_1 < 1 \; .$$

The first two moments of v_1 will now be obtained.

$$E v_1 = \frac{2^9 \cdot 7 \cdot 5}{3^8} \left[\int_0^1 (52v_1^8 - 178v_1^7 + 192v_1^6 - 29v_1^5 - 64v_1^4 \right.$$

$$+ 24v_1^3 + 4v_1^2 - v_1) \, dv_1$$

$$\left. + \int_0^{\frac{1}{4}} (v_1 - 4v_1^2 - 24v_1^3 + 64v_1^4 + 272v_1^5 - 192v_1^6 - 1280v_1^7 \right.$$

$$\left. - 1024v_1^8) \, dv_1 \right]$$

$$= \frac{2^9 \cdot 7 \cdot 5}{3^8} \left[\frac{52}{9} - \frac{178}{8} + \frac{192}{7} - \frac{29}{6} - \frac{64}{5} + \frac{24}{9} + \frac{4}{3} - \frac{1}{2} \right.$$

$$\left. + \frac{1}{2 \cdot 2^4} - \frac{4}{3 \cdot 2^6} - \frac{24}{4 \cdot 2^8} + \frac{64}{5 \cdot 2^{10}} + \frac{272}{6 \cdot 2^{12}} - \frac{192}{7 \cdot 2^{14}} \right.$$

$$- \frac{1280}{8 \cdot 2^{16}} - \frac{1024}{9 \cdot 2^{18}} \Big]$$

(4.3.20)

$$= .44342$$

$$E v_1^2 = \frac{2^9 \cdot 7 \cdot 5}{3^8} \Big[\int_0^1 (52v_1^9 - 178v_1^8 + 192v_1^7 - 29v_1^6 - 64v_1^5 + 24v_1^4 + 4v_1^3 - v_1^2)\, dv_1 + \int_0^{\frac{1}{4}} (v_1^2 - 24v_1^4 + 64v_1^5 + 272v_1^6 - 192v_1^7 - 1280v_1^2 - 1024v_1^9)\, dv_1 \Big]$$

$$= \frac{2^9 \cdot 7 \cdot 5}{3^8} \Big[\frac{52}{10} - \frac{178}{9} + \frac{192}{8} - \frac{29}{7} - \frac{64}{6} + \frac{24}{5} + \frac{4}{4} - \frac{1}{3} + \frac{1}{3 \cdot 2^6} - \frac{4}{4 \cdot 2^8} - \frac{24}{5 \cdot 2^{10}} + \frac{64}{6 \cdot 2^{12}} + \frac{272}{7 \cdot 2^{14}} - \frac{192}{8 \cdot 2^{16}} - \frac{1280}{9 \cdot 2^{18}} - \frac{1024}{10 \cdot 2^{20}} \Big]$$

(4.3.21)

$$= .21836 .$$

Hence

$$\mathrm{var}(v_1) = .02175 .$$

Since $u_1 = \frac{2v_1 + 1}{3}$ one obtains

(4.3.22) $$E u_1 = .62895$$

and

(4.3.23) $$E u_1^2 = E\left(\frac{4v_1^2 + 4v_1 + 1}{9} \right) = .40524 .$$

Now

$$\frac{u_3}{\sigma^2} = \frac{\ell_1 + \ell_2 + \ell_3}{\sigma^2} \sim x^2(18)$$

so that

(4.3.24) $$E u_3 = 18\sigma^2, \ \mathrm{var}(u_3) = 36\sigma^4, \text{ and } E u_3^2 = 360\sigma^4 .$$

Thus

(4.3.25) $$E\ \ell_1 = E\left[\frac{\ell_1}{\ell_1 + \ell_2 + \ell_3} \cdot (\ell_1 + \ell_2 + \ell_3)\right]$$

$$= E[u_1 u_3] = E u_1\ E u_3$$

$$= (.62895)\,18\sigma^2 = 11.3211\sigma^4$$

$$E\ \ell_1^2 = E u_1^2\ E u_3^2 = (.40524)(360\sigma^4)$$

$$= 145.8864\sigma^4$$

which implies

(4.3.26) $$\mathrm{var}(\ell_1) = 17.71909\sigma^4$$

and

(4.3.27) $$\text{S.D.}(\ell_1) = 4.20940\sigma^2 .$$

Now the upper $\alpha \cdot 100\%$ critical point for the distribution of v_1 when $b = 4$ and $t = 7$ will be obtained.

$$1 - \alpha = P_{Ho}\{v_1 < x\} = \int_0^x h_1(v_1)\,dx$$

$$= \frac{2^9 \cdot 7 \cdot 5}{3^8} \left[\int_0^{\frac{1}{4}} (1 - 4v_1 - 24v_1^2 + 64v_1^3 + 272v_1^4 - 192v_1^5 \right.$$

$$\left. - 1280v_1^6 - 1024v_1^7)\, dv_1 + \int_0^x (52v_1^7 - 178v_1^6 + 192v_1^5 \right.$$

$$\left. - 29v_1^4 - 64v_1^3 + 24v_1^2 + 4v_1 - 1)\, dv_1 \right]$$

unless the first integral in the bracket above is larger than

$\frac{3^8}{2^9 \cdot 7 \cdot 5}(1 - \alpha) = \frac{6561}{17920}(1 - \alpha)$. Thus

$$1 - \alpha = \frac{2^9 \cdot 7 \cdot 5}{3^8} \left(\frac{1}{2^2} - \frac{4}{2 \cdot 2^4} - \frac{24}{3 \cdot 2^6} + \frac{64}{4 \cdot 2^8} + \frac{272}{5 \cdot 2^{10}} - \frac{192}{6 \cdot 2^{12}} \right.$$

$$\left. - \frac{1280}{7 \cdot 2^{14}} - \frac{1024}{8 \cdot 2^{16}} + \frac{52x^8}{8} - \frac{178x^7}{7} + \frac{192x^6}{6} - \frac{29x^5}{5} - \frac{64x^4}{4} \right.$$

$$\left. + \frac{24x^3}{3} + \frac{4x^2}{2} - x \right)$$

$$= \frac{17920}{6561} \left(\frac{1697}{17920} + \frac{13}{2}x^8 - \frac{178}{7}x^7 + 32x^6 - \frac{29}{5}x^5 - 16x^4 \right.$$

$$\left. + 8x^3 + 2x^2 - x \right)$$

$$= \frac{17920}{6561} \left(\frac{1697}{17920} + (455x^8 - 1780x^7 + 2240x^6 - 406x^5 - 1120x^4 \right.$$

$$\left. + 560x^3 + 140x^2 - 70x) \div 70 \right).$$

Hence one has

$$455x^8 - 1780x^7 + 2240x^6 - 406x^5 - 1120x^4 + 560x^3 + 140x^2 - 70x$$

$+ \frac{1697 - 656(1 - \alpha)}{17920} = 0$ if $(1-\alpha) > \frac{1697}{6561}$. Solving the above equation for x when $\alpha = .10, .05, .01$ one obtains

$$(4.3.28) \quad x_{.10} = .64471\,,\ x_{.05} = .70051\,,\ x_{.01} = .79275$$

as the upper critical points when $b = 4$ and $t = 7$.

4.4 An Approximation to the Distribution of $\ell_1/(\ell_1 + \ell_2 + \cdots + \ell_{b-1})$

From (2.4.14) one can see that the density of

$u_1 = \dfrac{\ell_1}{\ell_1 + \ell_2 + \cdots + \ell_{b-1}}$ is nonzero only for $\frac{1}{b-1} < u_1 < 1$.

Consider the statistic

$$(4.4.1) \qquad v_1 = \frac{(b-1)u_1 - 1}{b-2} \quad \text{so that} \quad u_1 = \frac{(b-2)v_1 + 1}{b-1} \ .$$

Now

$$\frac{1}{b-1} < u_1 < 1$$

implies

$$\frac{1}{b-1} < \frac{(b-2)v_1 + 1}{b-1} < 1$$

which implies

$$1 < (b-2)v_1 + 1 < b - 1$$

which implies

$$0 < (b-2)v_1 < b - 2$$

which implies

(4.4.2) $$0 < v_1 < 1 .$$

Thus v_1 is nonzero only for $0 < v_1 < 1$. This suggests that one might consider approximating the distribution of v_1 with the Beta distribution.

Let $\mu_1 = E v_1$ and $\mu_2' = E v_1^2$. Setting these equal to the corresponding moments of the $\beta(a', b')$ distribution one has:

(4.4.3) $$\mu_1 = \frac{a'}{a' + b'} \quad \text{and} \quad \mu_2' = \frac{a'(a' + 1)}{(a' + b')(a' + b' + 1)} .$$

Solving for a' and b' one obtains

(4.4.4) $$a' = \frac{\mu_1(\mu_1 - \mu_2')}{\mu_2' - \mu_1^2} \quad \text{and} \quad b' = \frac{(1 - \mu_1)(\mu_1 - \mu_2')}{\mu_2' - \mu_1^2} .$$

Let x_1 be the point satisfying

(4.4.5) $$\int_0^{x_1} h_1(v_1)dv_1 = 1 - \alpha$$

where $h_1(v_1)$ is the density of

$$v_1 = \frac{(b - 1)u_1 - 1}{b - 2} = \frac{(b - 2)\ell_1 - (\ell_2 + \ell_3 + \cdots + \ell_{b-1})}{(b - 2)(\ell_1 + \ell_2 + \cdots + \ell_{b-1})} .$$

Let x_2 be the point satisfying

$$(4.4.6) \qquad \int_0^{x_2} \frac{1}{\beta(a', b')} w^{a'-1}(1-w)^{b'-1} dw = 1 - \alpha$$

where a' and b' are given in (4.4.4). Let

$$(4.4.7) \qquad \alpha^* = 1 - \int_0^{x_2} h_1(v_1) dv_1 \ .$$

Table 4.4.1 contains the values of x_1, x_2 and α^* for some selected values of α, b, and t. Observe from table 4.4.1 that the critical points obtained from the Beta approximation are quite close to the exact critical points. Even more important observe that α^* is not appreciably different than α.

If one had the first two moments of either u_1, v_1 or ℓ_1, the first two moments of v_1 could be obtained and thus the Beta approximation could be used. However it is not an easy problem to find the first two moments of any of these statistics in general. Mandel (1969) estimated the mean and variance of ℓ_1 by Monte-Carlo techniques for b = 4(1)8, 10, 12, 16, 20 and t = 3(1)8, 10, 12, 16, 20, 32, 50, 100. The values for b = 3, t = 3(1)8, 10, 12, 16, 20, 32, 50, 100. were also obtained but not published. In Table 4.4.2 the mean and variance of ℓ_1 as estimated by Mandel are compared with the mean and variance of ℓ_1 which has been found exactly for some selected values of b and t.

Now the moments of v_1 used in the Beta approximation will be given in terms of the moments of ℓ_1. Let $\ell_1 =$

$$(4.4.8) \qquad \upsilon_1 = \frac{E\,\ell_1}{\sigma^2} \quad \text{and} \quad \upsilon_2 = \frac{\text{var}(\ell_1)}{\sigma^4} \ .$$

Table 4.4.1

	$\alpha = .10$			$\alpha = .05$			$\alpha = .01$		
	x_1	x_2	α^*	x_1	x_2	α^*	x_1	x_2	α^*
b = 3 t = 3	.99499	.99394	.10994	.99875	.99818	.06038	.99995	.99983	.01858
t = 4	.94868	.94868	.10000	.97468	.97468	.05000	.99499	.99499	.01000
t = 5	.88575	.88605	.09963	.92967	.93012	.04953	.97652	.97685	.00979
t = 6	.82691	.82803	.09883	.88113	.88287	.04863	.94868	.95043	.00935
t = 7	.77582	.77763	.09824	.83564	.83873	.04788	.91734	.92117	.00893
t = 8	.73201	.73420	.09794	.79473	.79888	.04735	.88575	.89177	.00858
t = 9	.69430	.69653	.09792	.75836	.76318	.04704	.85541	.86337	.00833
t = 10	.66156	.66353	.09815	.72604	.73115	.04692	.82691	.83642	.00814
t = 12	.60749	.61045	.09717	.67136	.67795	.04608	.77582	.78828	.00778
t = 15	.54616	.54838	.09778	.60768	.61418	.04604	.71247	.72703	.00754
t = 20	.47512	.47636	.09864	.53210	.53810	.04611	.63286	.64904	.00728
t = 30	.38943	.38995	.09934	.43890	.44434	.04599	.52945	.54671	.00693
t = 50	.30246	.30274	.09956	.34256	.34751	.04556	.41785	.43504	.00650
b = 4 t = 5	.76872	.76117	.10884	.82173	.81086	.05907	.89550	.88587	.01474
b = 4 t = 7	.64471	.64050	.10480	.70051	.69303	.05558	.79275	.78112	.01307

Table 4.4.2

b	t	Exact Mean[1]	Estimated Mean[1]	Exact Variance[2]	Estimated Variance[2]
3	3	3.57	3.55	6.68	7.18
3	4	5.00	5.03	9.00	9.49
3	5	6.36	6.34	11.16	9.99
3	6	7.33	7.53	13.22	12.67
3	7	8.94	8.96	15.21	15.37
3	8	10.20	10.15	17.16	16.08
3	10	12.66	12.77	20.94	21.53
3	12	15.06	15.09	24.62	25.00
3	16	19.77	19.58	31.76	31.92
3	20	24.39	24.10	38.71	36.60
3	32	37.92	37.88	58.93	52.13
3	50	57.73	57.85	88.26	82.26
3	100	111.44	111.22	167.13	168.48
4	5	8.33	8.33	13.67	14.67
4	7	11.32	11.32	17.72	19.01

[1] Each value should be multiplied by σ^2.

[2] Each value should be multiplied by σ^4.

Now

$$E\,\ell_1 = E\,u_1\,E\,u_{b-1}$$

which implies

$$E\,u_1 = \frac{v_1}{(b-1)(t-1)}$$

and since

$$y_1 = \frac{(b-1)u_1 - 1}{b-2}$$

one obtains

$$(4.4.9) \qquad \mu_1 = E\,y_1 = \frac{\frac{v_1}{t-1} - 1}{b-2} = \frac{v_1 - (t-1)}{(b-2)(t-1)} \quad .$$

Also

$$E\,\ell_1^2 = E\,u_1^2\,E\,u_{b-1}^2$$

which implies

$$E\,u_1^2 = \frac{E\,\ell_1^2}{E\,u_{b-1}^2} = \frac{v_2 + v_1^2}{2(b-1)(t-1) + (b-1)^2(t-1)^2} = \frac{v_2 + v_1^2}{(b-1)(t-1)(bt-b-t+3)} \quad .$$

Since

$$y_1^2 = \frac{(b-1)^2 u_1^2 - 2(b-1)u_1 + 1}{(b-2)^2}$$

one obtains

(4.4.10)

$$\mu'_2 = E\, v_1^2 = \frac{\dfrac{(b-1)^2(v_2 + v_1^2)}{(b-1)(t-1)(bt-b-t+3)} - \dfrac{2(b-1)\,v_1}{(b-1)(t-1)} + 1}{(b-2)^2}$$

$$= \frac{(b-1)(v_2 + v_1^2) - 2(bt-b-t+3)v_1 + (t-1)(bt-b-t+3)}{(b-2)^2(t-1)(bt-b-t+3)} .$$

The mean and variance of ℓ_1 as estimated by Mandel are given in Appendix 1 .

4.5 An Estimate of σ^2 for Model (3.1.1)

If $\lambda = 0$ in the model (3.1.1), then $\dfrac{\ell_2 + \ell_3 + \cdots + \ell_{b-1}}{(b-1)(t-1) - v_1}$ is an unbiased estimate of σ^2 where v_1 was defined in (4.4.8). If $\lambda \neq 0$ in model (3.1.1), this writer is of the opinion that the value of ℓ_1 has little effect on the values of $\ell_2, \ell_3, \cdots, \ell_{b-1}$. Thus

(4.5.1) $$\frac{\ell_2 + \ell_3 + \cdots + \ell_{b-1}}{(b-1)(t-1) - v_1}$$

is a "good" estimate of σ^2 although if $\lambda \neq 0$, this estimate is no longer unbiased. Mandel (1969) supported this opinion with Monte Carlo Techniques. Note that the estimate (4.5.1) is a scale multiple of the maximum likelihood estimate for σ^2 found in Theorem 3.2.1.

4.6 An Example

In order to illustrate some of the ideas presented in the preceding sections, consider the following example. The data are the number of animal days of grazing for each calendar month on pastures located at Central Plains Experimental Range. It is a small part of the data obtained to determine the long-term effects of seasonal overgrazing. Permission to use the data was given by the Forage and Range Research Branch of the Agricultural Research Service.

Table 4.6.1 Animal days of grazing.

Years	Months												
	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	Totals
1966-67	144	112	130	140	116	108	128	63	84	88	88	96	1297
1967-68	155	126	125	270	245	255	208	210	280	250	230	220	2574
1968-69	240	168	204	168	130	112	92	88	84	88	60	92	1526
Totals	539	406	459	578	491	475	428	361	448	426	378	408	5397

The analysis of variance of the data in Table 4.6.1 is given in Table 4.6.2 with the single degree of freedom for non-additivity being that proposed by Tukey (1949).

Table 4.6.2 Analysis of Variance

SV	df	SS	MS	
Total	35	149,132.86		
Years	2	77,263.20	38,631.60	
Months	11	14,853.45	1,350.31	
Non-additivity	1	2,262.95	2,262.95	0.87
Residual	21	54,753.26	2,607.30	

By Tukey's single degree of freedom test for non-additivity, one must conclude that the data in Table 4.6.1 are additive.

Now consider the test proposed in (3.3.4). From the data in Table 4.6.1 one obtains $\ell_1 = 54,578.82$ and $\ell_2 = 2,437.39$. Thus

$\frac{\ell_1}{\ell_1 + \ell_2} = 0.957$. From Table 4.4.1 the 1% critical points is 0.77582. Thus the test proposed in (3.3.4) shows non-additivity significant at the 1% level while Tukey's Test showed no non-additivity.

From Table 4.4.2 the expected value of ℓ_1 is $15.06\sigma^2$. Thus the estimate of σ^2 proposed in (4.5.1) is

(4.6.1) $$\hat{\sigma}^2 = \frac{2437.39}{22 - 15.06} = \frac{2437.39}{6.94} = 351.21 \ .$$

In conclusion, this author would like to conjecture that the test proposed in (3.3.4) is a "better" test than Tukey's test for showing significant non-additivity in the following situations:

(1) The non-additivity in each cell is not a function of treatments and blocks.

(2) The data are non-additive because of a single outlier.

(3) There is no substantial difference in treatment means.

(4) There is no substantial difference in block means.

(5) The data are non-additive because only a small percentage of the total number of treatments or blocks are causing the non-additivity.

CHAPTER 5

PROBLEMS YET TO BE SOLVED

One of the main objectives of this paper was to find "good" estimates of σ^2 in the two-way classification model with one observation per cell. In Chapter 2 one solution to this problem was proposed. In determining the significance of the two by two table differences a critical point was used which was a function of $\ell_2 + \ell_3 + \cdots + \ell_{b-1}$ and was based on the distribution of $\ell_1/(\ell_2 + \ell_3 + \cdots + \ell_{b-1})$. In some models it might be desirable to determine the significance of the two by two table differences on a critical point which is a function of one of $\ell_3 + \ell_4 + \cdots + \ell_{b-1}$, $\ell_4 + \ell_5 + \cdots + \ell_{b-1}, \cdots, \ell_{b-1}$ and based on the distribution of one of $\ell_1/(\ell_3 + \ell_4 + \cdots + \ell_{b-1})$, $\ell_1/(\ell_4 + \ell_5 + \cdots + \ell_{b-1}), \cdots, \ell_1/\ell_{b-1}$. Thus one would be interested in knowing how many of the characteristic roots $\ell_1, \ell_2, \cdots, \ell_{b-1}$ are accounting for something in the residual sum of squares other than error. If one decided that the roots $\ell_k, \ell_{k+1}, \cdots, \ell_{b-1}$; $k \geq 2$ are accounting for nothing more than error, then one would want the distribution of $\ell_1/(\ell_k + \ell_{k+1} + \cdots + \ell_{b-1})$ under the hypothesis that all two by two table differences are zero. In this case it would appear that this would be a more powerful test for the two by two table differences than the test based on $\ell_1/(\ell_2 + \ell_3 + \cdots + \ell_{b-1})$.

Another improvement to the methods in Chapter 2 would be to find the distribution of Q or to find the distribution of an upper bound closer to Q than ℓ_1.

Another problem to be considered is how well can one interpret the pattern of interaction in the two-way classification model by using two by two table differences if one had an independent estimate of σ^2. This independent estimate could be obtained from some other source or a replicated experiment. One almost obvious solution to the

significance of two by two table differences in this case is to use the multiple comparison techniques in Scheffe (1959), known to some as the S-method. However, it seems that one should be able to improve upon the S-method since one knows the types of contrasts of the observations he is interested in using.

Another problem that comes to mind is extending the two by two table differences method to other designs such as the n-way classification designs, latin square designs, Balanced Incomplete Block designs, and perhaps eventually to the general linear model.

There was no attempt made in this paper to interpret the information available in the characteristic vectors of $Z'Z$ and ZZ' corresponding to the largest root ℓ_1. In large experiments the number of two by two table differences is staggering and the characteristic vectors could be very useful. As an example suppose the model is (3.1.1). If $\alpha_i = \alpha'_i$, then every two by two table difference involving treatments i and i' will be zero. Thus one could obtain an estimate of σ^2 from those two treatments.

One would also be interested in the non-null distribution of $\ell_1/(\ell_1 + \ell_2 + \cdots + \ell_{b-1})$ which could be used to examine the power of the likelihood ratio test of $Ho: \lambda = 0$ in model (3.1.1). If one knew the moments of $\ell_2 + \ell_3 + \cdots + \ell_{b-1}$ under $Ha: \lambda \neq 0$, one would have a better idea of the appropriateness of the estimate of σ^2 proposed in (4.5.1).

It has also been conjectured that the likelihood ratio test of $Ho: \lambda = 0$ in model (3.1.1) has many of the same properties as the usual F-test in the analysis of variance. At this time no specific conclusions have been drawn.

One would also be interested in a set of minimal sufficient statistics for the model (3.1.1).

The unsolved problems discussed above that are being considered are summarized below..

(1) Determine how many of the characteristic roots of $Z'Z$ are accounting for a significant amount of the interaction.

(2) Consider the use of some other functions of the roots of $Z'Z$ in order to test for the significance of two by two table differences.

(3) Consider the model

$$y_{ij} = \mu + \tau_i + \beta_j + \lambda_1 \alpha_{1i} \gamma_{1j} + \lambda_2 \alpha_{2i} \gamma_{2j} + \cdots$$

as an alternative to the additive model.

(4) Find the distribution of Q or the distribution of an upper bound of Q which is closer to Q than ℓ_1.

(5) Consider the model

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + e_{ijk} \quad i = 1, 2, \cdots, t; \; j = 1, 2, \cdots, b;$$

$k = 1, 2, \cdots, r (> 1)$, and perfect a technique of using two by two table differences to discover the pattern of interaction.

(6) Extend the method of two by two table differences to other designs.

(7) Consider the problem of extracting the information that is available in the characteristic vectors of $Z'Z$ and ZZ'.

(8) Consider the non-null distribution of some useful functions of the characteristic roots of a matrix having the noncentral Wishart distribution.

(9) Determine a set of minimal sufficient statistics for the model (3.1.1).

(10) Consider the properties of the likelihood ratio test for non-additivity in the model (3.1.1) and study the power of this test.

(11) Get percentage points of $\ell_1/(\ell_1 + \ell_2 + \cdots + \ell_{b-1})$ for all values of b and t.

In conclusion, the author feels that the methods proposed in this paper can be very valuable in the interpretation and the analysis of two-way classification models.

APPENDIX I

Table A1. Expected values of ℓ_1/σ^2 from Mandel (1969).

t \ b	4	5	6	7	8	10	12	16	20
4	6.45	8.47	9.86	11.61	12.88	15.08	17.96	23.33	28.20
5	8.47	10.37	11.82	13.59	14.75	18.11	21.01	26.29	31.87
6	9.86	11.82	13.35	15.44	16.92	20.42	22.85	28.87	36.78
7	11.61	13.59	15.44	17.18	18.91	22.89	25.59	31.83	37.46
8	12.88	14.75	16.92	18.91	20.72	24.15	27.67	34.52	40.70
10	15.08	18.11	20.42	22.89	24.15	27.81	31.88	38.81	45.11
12	17.96	21.01	22.85	25.59	27.67	31.88	35.39	42.76	50.00
16	23.33	26.29	28.87	31.83	34.52	38.81	42.76	51.08	58.11
20	28.20	31.87	36.78	37.46	40.70	45.11	50.00	58.11	66.34
32	42.86	46.77	50.83	53.77	57.50	63.21	68.60	78.62	88.71
50	63.85	68.93	73.74	77.62	81.68	88.66	95.11	107.41	117.36
100	120.90	126.58	132.94	138.45	143.64	153.00	161.58	176.32	192.08

Table A2. Standard deviation of ℓ_1/σ^2 from Mandel (1969).

t \ b	4	5	6	7	8	10	12	16	20
4	3.24	3.83	4.05	4.36	4.52	4.72	5.26	5.71	6.25
5	3.83	4.18	4.12	4.28	4.71	5.12	5.45	5.84	6.40
6	4.05	4.12	4.30	4.88	4.71	5.64	5.51	6.09	6.92
7	4.36	4.28	4.88	4.80	5.16	5.28	5.64	6.03	6.51
8	4.52	4.71	4.71	5.16	5.26	5.30	6.02	6.49	6.88
10	4.72	5.12	5.64	5.28	5.30	5.50	6.41	6.25	6.89
12	5.26	5.45	5.51	5.64	6.02	6.41	6.65	6.57	7.53
16	5.71	5.84	6.09	6.03	6.49	6.25	6.57	7.10	7.60
20	6.25	6.40	6.92	6.51	6.88	6.89	7.53	7.60	7.87
32	7.38	7.59	7.96	7.89	7.98	8.02	8.26	8.37	8.77
50	9.55	9.55	9.44	8.88	9.04	9.36	9.77	9.86	10.16
100	12.96	12.22	12.50	11.84	11.96	11.67	11.59	11.95	12.19

BIBLIOGRAPHY

Abraham, J.K. (1960), "Note: On an alternative method of computing Tukey's statistic for the Latin Square Models", Biometrics, 16 : 686-691.

Anscombe, F.J. (1961), "Examination of residuals", 1 Proceedings of the Fourth Berkely Symposium on Mathematical Statistics and Probability (University of California Press).

Anscombe, F.J. and Tukey, John W. (1963), "The examination and analysis of residuals", Technometrics, 5 : 141-160.

Black, A.L. (1970), "Adventitious roots, tillers, and grain yields of spring wheat as influenced by N - P fertilization", Agronomy Journal 62 : 32-36.

Elston, R.C. (1961), "On additivity in the analysis of variance", Biometrics, 17 : 209-219.

Elston, R.C. and Bush, N. (1964), "The hypotheses that can be tested when there are interactions in an analysis of variance model", Biometrics, 20 : 681-698.

Fisher, R.A. (1939), "The sampling distribution of some statistics obtained from non-linear equations", Annals of Eugenics, 9 : 238-249.

Ghosh, M.N and Sharma, Divakar (1963), "Power of Tukey's test for non-additivity", Journal of the Royal Statistical Society, B, 25 : 213-219.

Gilbert, Neil (1963), "Non-additive combining abilities", Genetic Research, Cambridge, 4 : 65-73.

Girshick, M.A. (1939), "On the sampling theory of roots of determinantal equations", Annals of Mathematical Statistics, 10 : 203-224.

Guanadesikan, R. and Wilk, M.B. (1970), "Use of maximum likelihood for estimating error variance from a collection of analysis of variance mean squares", Annals of Mathematical Statistics, 41 : 292-304.

Gollob, H.F. (1968), "A statistical model which combines features of factor analytic and analysis of variance techniques", Psychometrika, 33 : 73-116.

Graybill, F. A. (1961), An Introduction to Linear Statistical Models, Vol. I, McGraw-Hill: New York.

Graybill, F. A. (1969), Introduction to Matrices with Applications in Statistics, Wadsworth: Belmont, California.

Graybill, F. A. and Milliken, G. A. (1969), "Quadratic forms and idempotent matrices with random elements", Annals of Mathematical Statistics, 40: 1430-1438.

Harter, H. Leon and Lum, Mary D. (1962), An interpretation and extension of Tukey's one degree of freedom for non-additivity, Aeronautical Research Laboratory, Wright Patterson Air Force Base, Ohio. (ARL-62-313).

Hayakawa, T. (1967), "On the distribution of the maximum latent root of a positive definite symmetric random matrix", Annals of the Institute of Statistical Mathematics, 20: 1-17.

Hsu, P. L. (1939), "On the distribution of the roots of certain determinantal equations", Annals of Eugenics, 9: 250-258.

James, A. T. (1961), "The distribution of noncentral means with known covariance", Annals of Mathematical Statistics, 32: 874-882.

James, A. T. (1964), "Distributions of matrix variates and latent roots derived from normal samples", Annals of Mathematical Statistics, 35: 475-501.

Khatri, C. G. and Pillai, K. C. S. (1968), "On the noncentral distribution of two test criterion in multivariate analysis of variance", Annals of Mathematical Statistics, 39: 215-226.

Mandel, John (1959), "The analysis of latin squares with a certain type of row-column interaction", Technometrics, 1: 379-387.

Mandel, John (1961), "Non-additivity in two-way analysis of variance", Journal of the American Statistical Association, 56: 878-888.

Mandel, John (1964), "Estimation of weighting factors in linear regression and analysis of variance", Technometrics, 6: 1-25.

Mandel, John (1969), "A method for fitting emperical surfaces to physical or chemical data", Technometrics, 11: 411-429.

Milliken, G. A. (1970), Extensions of the General Linear Model, Unpublished Ph. D. Thesis, Colorado State University.

Milliken, G. A. and Graybill, F. A. (1970), "Extensions of the general linear hypothesis model", Journal of the American Statistical Association, 65 : 797-807 .

Mood, A. M. (1951), "On the distribution of the characteristic roots of normal second-moment matrices", Annals of Mathematical Statistics, 22 : 266-273 .

Penrose, R. (1955), "A generalized inverse for matrices", Proc. Camb. Phil. Soc. 51 : 406-413 .

Pillai, K. C. S. (1956), "On the distribution of the largest or the smallest root of a matrix in multivariate analysis", Biometrika, 43 : 122-127 .

Pillai, K. C. S. (1967), "On the distribution of the largest root of a matrix in multivariate analysis", Annals of Mathematical Statistics, 38 : 616-617 .

Rao, C. R. (1965), Linear Statistical Inference and Its Applications, Wiley : New York .

Roy, S. N. (1939), "p-statistics, or some generalizations in analysis of variance appropriate to multivariate problems", Sankhya , 4 : 381-396 .

Scheffe, H. (1959), The Analysis of Variance, Wiley : New York .

Seber, G. A. F. (1966), The Linear Hypothesis, Hefner : New York .

Sugiyama, T. (1966), "On the distribution of the largest latent root and the corresponding latent vector for principle component analysis", Annals of Mathematical Statistics, 37 : 995-1001 .

Sugiyama, T. (1967), "On the distribution of the largest characteristic root of the covariance matrix", Annals of Mathematical Statistics, 38 : 1148-1151 .

Tukey, John W. (1949), "One degree of freedom for non-additivity", Biometrics, 5 : 232-242 .

Tukey, John W. (1955), "Answer to query 113", Biometrics, 11 : 111-113 .

Tukey, John W. (1962), "The future of data analysis", Annals of Mathematical Statistics, 33 : 1-67 .

van der Vaart, H. R. (1961), "On certain characteristics of the distribution of a symmetric random matrix under general conditions", Annals of Mathematical Statistics, 32 : 864-873.

Ward, G. C. and Dick, I. D. (1952), "Non-additivity in randomized block designs and balanced incomplete block designs", New Zealand Journal of Science and Technology, 33 : 430-435.

Wilk, M. B., Gnandesikan, R. and Freeny, Anne E. (1963), "Estimation of error variance from smallest ordered contrasts", Journal of the American Statistical Association, 58 : 152-160.

Williams, E. J. (1952), "The interpretation of interactions in factorial experiments", Biometrika, 39 : 65-81.