The non-orthogonal analysis of variance

Richard LaVern Chamberlain

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

https://lib.dr.iastate.edu/rtd/5205

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company
CHAMBERLAIN, Richard LaVern, 1942-
THE NON-ORTHOGONAL ANALYSIS OF VARIANCE.
Iowa State University, Ph.D., 1972
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan
The non-orthogonal analysis of variance

by

Richard LaVern Chamberlain

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1972
PLEASE NOTE:

Some pages may have
indistinct print.
Filmed as received.

University Microfilms, A Xerox Education Company
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II.</td>
<td>LITERATURE REVIEW</td>
<td>6</td>
</tr>
<tr>
<td>A.</td>
<td>Orthogonality and Non-orthogonality</td>
<td>15</td>
</tr>
<tr>
<td>B.</td>
<td>Models of Maximal Rank</td>
<td>17</td>
</tr>
<tr>
<td>C.</td>
<td>Models of Submaximal Rank</td>
<td>20</td>
</tr>
<tr>
<td>D.</td>
<td>An Aspect of Reparameterization</td>
<td>22</td>
</tr>
<tr>
<td>E.</td>
<td>Computational Procedures</td>
<td>24</td>
</tr>
<tr>
<td>1.</td>
<td>Wilkinson's algorithm</td>
<td>24</td>
</tr>
<tr>
<td>2.</td>
<td>Fowlkes' algorithm</td>
<td>44</td>
</tr>
<tr>
<td>3.</td>
<td>Bradley's algorithm</td>
<td>46</td>
</tr>
<tr>
<td>III.</td>
<td>IMPOSING ESTIMABLE CONDITIONS</td>
<td>52</td>
</tr>
<tr>
<td>A.</td>
<td>Least Squares Derivation</td>
<td>52</td>
</tr>
<tr>
<td>B.</td>
<td>Degrees of Freedom, Sums of Squares, and Tests of Hypotheses</td>
<td>56</td>
</tr>
<tr>
<td>C.</td>
<td>The Variance of Linearly Estimable Functions</td>
<td>57</td>
</tr>
<tr>
<td>D.</td>
<td>The Special Case When ( c = 0 )</td>
<td>58</td>
</tr>
<tr>
<td>IV.</td>
<td>PARTITIONED MODELS</td>
<td>61</td>
</tr>
<tr>
<td>A.</td>
<td>The Linear Model With Two Partitions</td>
<td>62</td>
</tr>
<tr>
<td>B.</td>
<td>Estimability</td>
<td>63</td>
</tr>
<tr>
<td>C.</td>
<td>Estimation in Submodels</td>
<td>74</td>
</tr>
<tr>
<td>D.</td>
<td>Decomposition of Sums of Squares</td>
<td>78</td>
</tr>
<tr>
<td>E.</td>
<td>The Reduced Normal Equations</td>
<td>96</td>
</tr>
<tr>
<td>F.</td>
<td>The Variances of Estimates of Estimable Functions</td>
<td>104</td>
</tr>
<tr>
<td>G.</td>
<td>Interfactor Information</td>
<td>108</td>
</tr>
</tbody>
</table>
### V. THE TWO-WAY FACTORIAL

**A. Without Interaction**
- The balanced complete case
- The general case

**B. With Interaction**
- The balanced complete case
- The general case

### VI. NESTED FACTORS

**A. Estimation**
- The balanced complete case
- The general case

**B. Tests of Hypotheses**

### VII. MULTIFACTOR EXPERIMENTS

**A. Estimation**

**B. The Maximum Number of Partitions**

**C. Orthogonal Data Situations**

**D. Non-orthogonal Data Situations**
- The geometry of the non-orthogonal analysis of variance
- Partial orthogonality

**E. Decomposition of Sums of Squares**

**F. Relationship to More Standard Analyses.**

**G. The Variances of BLUE's of Estimable Functions**

### VIII. GENERALIZED LEAST SQUARES
<table>
<thead>
<tr>
<th>IX.</th>
<th>A COMPUTATIONAL PROCEDURE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Analysis of Variance Programs</td>
<td>210</td>
</tr>
<tr>
<td>B.</td>
<td>Constructing a Basis for the Intersection of Two Column Spaces</td>
<td>212</td>
</tr>
<tr>
<td>C.</td>
<td>The Two-factor Model</td>
<td>217</td>
</tr>
<tr>
<td>D.</td>
<td>Models With More Than Two Factors</td>
<td>221</td>
</tr>
<tr>
<td>X.</td>
<td>SUMMARY</td>
<td>224</td>
</tr>
<tr>
<td>A.</td>
<td>Estimability</td>
<td>224</td>
</tr>
<tr>
<td>B.</td>
<td>Comparison With More Classical Procedures</td>
<td>225</td>
</tr>
<tr>
<td>C.</td>
<td>Further Work</td>
<td>225</td>
</tr>
<tr>
<td>XI.</td>
<td>ACKNOWLEDGEMENTS</td>
<td>227</td>
</tr>
<tr>
<td>XII.</td>
<td>BIBLIOGRAPHY</td>
<td>228</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

The development of the computer and its application to the field of statistics over the past twenty years has led to numerous computer algorithms for performing a wide variety of statistical analyses. Not the least of these has been for the statistical technique known as the analysis of variance. The degree of generality of these algorithms varies considerably. The types of input to them as well as the information generated by them seem to vary just as much. There is, of course, a wide variety of information that can be extracted from a set of data through the use of the analysis of variance. However, names assigned to various pieces of information sometimes differ from one author to another, and methods used to obtain one piece of information can make it difficult to obtain another piece of information. As an example of the latter, Wilkinson (1970) calculates essentially the minimal polynomial of a matrix and uses this to calculate a particular conditional inverse. Yet, the minimal polynomial of a matrix contains no real information concerning the rank of the matrix which is necessary for calculating the degrees of freedom in an analysis of variance.

Many statisticians, particularly practicing statisticians, are often faced with data that can be classified as non-orthogonal. If the data have a planned non-orthogonal structure, such as occurs with Incomplete Block Designs, there is a vast literature to which one can turn. However, when the data are non-orthogonal because of extenuating circumstances, poor planning, error in designing the experiment, or a natural disaster, it is not at all clear what computational processes are adequate and efficient.
Determining what information is contained in the experimental data is not, in general, a simple task; and extraction of this information may be difficult. When faced with these arbitrary experimental designs, the trained human mind has a capacity for recognizing structures that is difficult, if not impossible, to duplicate in a computer. Even distinguishing between a very small number and the number zero becomes a non-trivial task in the computer.

To obtain an idea of the various approaches to analysing an arbitrary experimental design, one need only read the report on the Analysis of Variance Workshop held at the University of Wisconsin which was prepared by Muller and Wilkinson (1971). The variety of forms for input and output is evident. The seeming inability of some of these procedures to handle even relatively simple non-orthogonal designs such as that presented by Meeks in the above report (page 39) indicates that there is still some work to be done in this area.

There has been a vast amount of study of the fitting of the linear model, usually represented by the matrix model equation, \( y = X\beta + \epsilon \), for various covariance structures of the error component, \( \epsilon \). There seems to be a less thorough treatment of the partitioned model represented by the matrix model equation, \( y = X_1\beta_1 + X_2\beta_2 + \epsilon \). Such a partitioned model has been used widely in connection with the statistical procedure known as the analysis of covariance, particularly before the advent of the modern computer. Such a model also is the basis for the statistical analysis associated with linear classificatory models. In the days of the desk calculator, the data analyst performed all the steps in a statistical analysis with comprehension and realization of what was being done at each
step. This is in contrast to modern computer technology in which the whole analysis of at least one pass through the data must be preplanned and the meaning of the output must be based on complete knowledge of the whole algorithm.

Another aspect of the problem is that while it is convenient for some purely mathematical approaches to use what is called a coordinate free approach, a particular coordinate known approach is usually required by the experimenter. The experimenter is not merely interested in fitting the model, say, \( y = X\beta + \varepsilon \). He is interested in aspects of the vector parameter \( \beta \) that are used in his particular coordinate representation. He is concerned with knowing the linear functions of the elements of \( \beta \) that are estimable, the estimates of these functions, their variances, and so on. A general computing algorithm must include answers to these problems.

Still another part of the problem of analysing an arbitrary experimental structure arises when the data and associated linear model do not satisfy the property known as maximal rank. A linear model and data situation is said to be of maximal rank if the data are sufficient to provide estimates of all estimable functions that would be estimable with an indefinitely large amount of data. The occurrence of submaximal rank and the consequent problems of determining what is estimable are the Achilles Heel of recent computing algorithms.

The problems mentioned above form the impetus for this thesis. The partitioned model \( y = X_1\beta_1 + X_2\beta_2 + \varepsilon \) is examined in some detail. More complex models of the form \( y = X_1\beta_1 + \ldots + X_p\beta_p + \varepsilon \) are introduced and the results for \( p = 2 \) are extended to arbitrary positive integer values.
of $p$. Finally, a computational procedure is proposed for analysing arbitrary experimental structures.

Chapter II contains a review of pertinent literature concerning non-orthogonal analysis of variance. Necessary definitions and notations are presented. Various degrees of non-orthogonality are discussed and illustrated. The last part of the chapter contains descriptions of several recent algorithms for analysing complex experimental designs. Of particular interest is the algorithm due to Wilkinson (1970).

Chapter III presents some useful results concerning the imposing of estimable conditions on the linear model,

$$y = X\beta + \epsilon.$$  

These results establish some notational conventions and are used in later chapters.

Chapter IV is one of the main chapters in the thesis. The theoretical results concerning estimation, decomposition of sums of squares, expected mean squares, and tests of hypothesis are presented for a general two-factor model. The framework is laid for the extension of these results to models with more than two factors.

Chapter V discusses a simple two-factor model with and without interaction. Some of the problems concerning the estimation and calculation of expected mean squares in the presence of interaction are explored. An example of a model of submaximal rank is presented and analysed in some detail.
Chapter VI is similar in nature to Chapter V, except that the model being discussed is a two-factor nested model. Some contingencies when unequal numbers and missing cells are present are discussed.

In Chapter VII some projection operators are defined that aid in deriving extensions of the results of previous chapters to multifactor models. Results concerning the maximum number of partitions of the estimation space for special models and the analysis of arbitrary experimental structures are presented.

Chapter VIII discusses the extension of results of previous chapters to the case where the errors have a general covariance structure. The results of Zyskind and Martin (1969) are used to derive generalized reduced normal equations and generalized reduced conjugate normal equations.

In Chapter IX some computational procedures are presented in the light of the results of the previous chapters. The computational aspects of a general computing algorithm are outlined.

Chapter X is a summary and a discussion of areas in which more work needs to be done.

This thesis does not answer all of the questions centered around the analysis of non-orthogonal data. One of the reasons for this is that many of these questions are not statistical in nature. They can only be answered by the person conducting the experiment and from his familiarity with the experimental material. However, it is the responsibility of the statistician to lay bare the statistical information in a set of data taking into account its source and the questions being asked.
II. LITERATURE REVIEW

The literature concerning non-orthogonal analysis of variance is vast. Kempthorne (1952), Scheffé (1959), and Barcroft (1968), to name only a few, discuss the analysis of non-orthogonal designs encountered when there are unequal numbers of observations taken at various levels of the factors involved in the experiment. The analysis of planned non-orthogonal designs such as Incomplete Block Designs is discussed by Yates (1936), Bose (1949), Kempthorne (1952), Bose and Mesner (1959), Shah (1959), Folks and Kempthorne (1960), and others too numerous to mention. Therefore, in order to keep the problem within focus, it will be necessary to limit the discussion here to a particular approach to non-orthogonal data. For this reason, the discussion does not concern itself, except indirectly, with the design of non-orthogonal experiments; instead, the discussion centers around the analysis of non-orthogonal data. Such data may arise from a planned experiment, but more often it arises from an observational study in which the occurrence of factor levels and combinations of factor levels do not have properties of balance which would be sought in a planned experimental study.

In order to clarify the exposition that follows, we will introduce some notational conventions and define some quantities that are used considerably. Of primary importance is the concept of a linear model, written in matrix notation as

\[(2.1) \quad y = X\beta + \varepsilon,\]
where \( y \) is an \( n \times 1 \) vector of observations, \( X \) is an \( n \times m \) matrix of constants (referred to as the model matrix), \( \beta \) is an \( m \times 1 \) matrix of parameters that we are generally trying to estimate, and, unless stated otherwise, \( \varepsilon \) is an \( n \times 1 \) vector of random variables with mean zero and variance covariance matrix equal to \( \sigma^2 I \). At times it will be more informative to write 2.1 in its partitioned form,

\[
(2.2) \quad y = X_1 \beta_1 + X_2 \beta_2 + \ldots + X_p \beta_p + \varepsilon ,
\]

where the model matrix \( X \) has been partitioned into \( X = (X_1, X_2, \ldots, X_p) \), and the parameter vector \( \beta' \) has been partitioned accordingly into \( \beta' = (\beta_1', \beta_2', \ldots, \beta_p') \). Each partition \( X_i \) is \( n \times p_i \) and each \( \beta_i \) is \( p_i \times 1 \) with

\[
(2.3) \quad \sum_{i} p_i = m .
\]

The additive linear model for a simple experiment involving two factors, say \( A \) and \( B \), is usually written as

\[
(2.4) \quad y_{ij} = \mu + a_i + b_j + \varepsilon_{ij} ,
\]

where \( \mu \) is a constant for all \( y_{ij} \), \( a_i \) corresponds to the effect produced by the \( i \)-th level of factor \( A \), \( i = 1, 2, \ldots, 4 \), \( b_j \) represents the effect produced by the \( j \)-th level of factor \( B \), \( j = 1, 2, \ldots, 4 \), and \( \varepsilon_{ij} \) is the experimental error associated with the \( i,j \)-th factor combination. Equation 2.4 can be written in the form of 2.2 by letting
1) \( p_1 = 1, p_2 = 5, p_3 = 4, n = 4t, \)

ii) \( \beta_1 = \mu, \beta_2 = \begin{bmatrix} a_1 \\ \vdots \\ a_4 \end{bmatrix}, \beta_3 = \begin{bmatrix} b_1 \\ \vdots \\ b_4 \end{bmatrix} \)

iii) \( y = \begin{bmatrix} y_{11} \\ \vdots \\ y_{4t} \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{4t} \end{bmatrix} \)

iv) \( X_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, X_2 = (x_{ki}^{(2)}), X_3 = (x_{kj}^{(3)}) \)

where \( x_{ki}^{(2)} = 1 \) if the \( k \)-th observation, \( k = 1, \ldots, n \), is associated with \( i \)-th level of factor \( A \) and \( x_{ki}^{(2)} = 0 \) otherwise; and \( x_{kj}^{(3)} = 1 \) if the \( k \)-th observation, \( k = 1, \ldots, n \), is associated with the \( j \)-th level of factor \( B \) and \( x_{kj}^{(3)} = 0 \) otherwise.

Consider, for example, the two-factor experiment where \( A \) has two levels and \( B \) has three levels. Expression 2.4 can be written
\[ y_{11} = \mu + a_1 + b_1 + \varepsilon_{11} \]
\[ y_{12} = \mu + a_1 + b_2 + \varepsilon_{12} \]
\[ y_{13} = \mu + a_1 + b_3 + \varepsilon_{13} \]
\[ y_{21} = \mu + a_2 + b_1 + \varepsilon_{21} \]
\[ y_{22} = \mu + a_2 + b_2 + \varepsilon_{22} \]
\[ y_{23} = \mu + a_2 + b_3 + \varepsilon_{23} . \]

(2.5)

The corresponding representation as in 2.2 is

\[ \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} \mu \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix} . \]

(2.6)

Models such as 2.6, where the model matrix is made up of 0's and 1's, are called linear classificatory models because of the way this induces a classification of the response into effects of levels of factors. The use of 0's and 1's in the model matrix is entirely arbitrary, other constants can be used just as well. Unless otherwise stated, the model matrices associated with linear classificatory models in this presentation will be made up of 0's and 1's.
Expression 2.6 can also be used to illustrate another point. Notice that the sum of the two columns of $X_2$ is the matrix $X_1$, also that the sum of the three columns of $X_3$ gives $X_1$. That is, the matrix $X_1$ is made up of a linear combination of the columns of $X_2$. Equivalently, $X_1$ is contained in the vector space generated by the columns of $X_2$. This space is called the column space of $X_2$ and is denoted by $C(X_2)$. Similarly, the matrix $X_1$ is contained in the column space of $X_3$, $C(X_3)$. It is also useful to define the row space of a matrix as the vector space generated by all linear combinations of the rows of the matrix. For example, the row space of $X_2$, denoted by $R(X_2)$, is the set of all pairs of real numbers, $(u,v)$.

The fact that $X_1$ is contained in both $C(X_2)$ and $C(X_3)$ suggests another convention used by some authors. Authors such as Potthoff (1962a, 1962b), when discussing linear classificatory models, do not include the mean $\mu$ in the model. Equation 2.4 is written as

$$y_{ij} = a_i + b_j + \varepsilon_{ij}. \tag{2.7}$$

Because of the containment property, $X_1 \in C(X_2)$ and $X_1 \in C(X_3)$, specification of the mean is, in a sense, redundant. The fact that the matrix $X = (X_1, X_2, X_3)$ is $n \times p$ of rank $p - 2$, implies that only certain linear functions of the parameters can be estimated. When a linear function of the parameters, $\lambda'\beta$, can be estimated, that linear function is said to be estimable. It is well known that such linear functions satisfy $\lambda' \in R(X)$.

Using the method of least squares to estimate the parameters, we find a vector $\hat{\beta}$ such that the function
\[(2.8) \quad Q(\beta) = (y-X\beta)'(y-X\beta)\]

is a minimum. It is well known that \(Q(\beta)\) is a minimum for any \(\hat{\beta}\) satisfying the normal equations

\[(2.9) \quad X'X\hat{\beta} = X'y .\]

If the matrix \(X\) is \(n \times p\) of rank \(p\), then the model 2.1 is said to be of full rank and the matrix \(X'X\) is nonsingular. In the case of a model of full rank, all elements of the parameter vector \(\beta\) are estimable, and their Best Linear Unbiased Estimates (BLUE\'s) are given by

\[(2.10) \quad \hat{\beta} = (X'X)^{-1}X'y ,\]

where best is used in the sense of minimum variance.

If the model is not of full rank, then there exists an infinite number of vectors \(\hat{\beta}\) that satisfy the normal equations. A common way of solving the normal equations is to impose conditions on a suitably chosen set of linear functions of the parameters. The augmented set of linear equations must be consistent and have a unique solution. Another procedure is to express the model as one of full rank on a complete set of linearly independent estimable functions. This is called reparameterization and results in a new vector of parameters, say \(\Theta\), and a new model matrix \(W\) such that \(C(W) = C(X)\) and the reparameterized model

\[(2.11) \quad y = W\Theta + \epsilon ,\]

is of full rank.
Another method for obtaining solutions to the normal equations when the matrix $X'X$ is singular is by means of generalized inverses or conditional inverses. The concept of a generalized inverse was developed independently by Moore (1935) and Penrose (1955). The generalized inverse of any real matrix $A$ denoted by $A^+$ satisfies the following conditions:

\begin{align}
\text{i)} & \quad AA^+A = A \\
\text{ii)} & \quad A^+AA^+ = A^+ \\
\text{iii)} & \quad (AA^+)' = AA^+ \\
\text{iv)} & \quad (A^+A)' = A^+A.
\end{align}

(2.12)

Any matrix $A^-$ that satisfies only condition i) is called a conditional inverse of $A$. In the case where $A$ is nonsingular,

\begin{align}
(2.13) & \quad A^+ = A^- = A^{-1}.
\end{align}

The matrix $A^+$ is unique for any matrix $A$; however, in general the matrix $A^-$ is not unique. A discussion of generalized and conditional inverses, linear models, and other concepts mentioned above can be found in Zyskind, Kempthorne, et al. (1964). Any vector $\hat{\beta}$ of the form

\[ \hat{\beta} = (X'X)^{-1}X'y \]

or

\[ \hat{\beta} = (X'X)^{+}X'y \]

satisfies the normal equations.
Linear classificatory models form the basis of many experimental studies. Every observation is classified according to each of a set of factors, say, \(F_1, F_2, \ldots, F_k\). Each observation is then indexed by a \(k\)-tuple denoting the level of each factor associated with the observation. An observation can be denoted by \(y(x_1, x_2, \ldots, x_k)\). (Sometimes the \(k\)-tuple is written as a subscript on the variable \(y\).) There are intrinsically two types of relations between factors. These are called crossing and nesting. We shall first discuss completely crossed structures.

For completely crossed structures, the indexing variables \(x_i\) can assume any of the values 1, 2, \ldots, \(L_i\), say, where \(L_i\) is the number of levels of factor \(F_i\). There is then a total of \(\prod_{i=1}^{k} L_i\) possible combinations. If every possible combination occurs an equal number of times in the total set of data, fitting a standard linear model is straightforward, although there are some unresolved inferential problems. One can examine the data by means of a full factorial model of the form

\[
E[y(x_1, \ldots, x_k)] = \mu + f^{(1)}_{x_1} + f^{(2)}_{x_2} + \ldots + f^{(1,2)}_{x_1 x_2} + \ldots + f^{(1,2,\ldots,k)}_{x_1 x_2 \ldots x_k},
\]

where \(\mu\) is a number which is constant for all observations, \(f^{(1)}_{x_1}\) is a number associated with all observations at the \(x_1\)-th level of factor \(F_1\), \(f^{(1,2)}_{x_1 x_2}\) is a number associated with all combinations at the \(x_1\)-th level of factor \(F_1\) and the \(x_2\)-th level of factor \(F_2\), and so on. Terms involving more than one indexing variable are called interactions. One can also examine the data with a model obtained by deleting all of the
terms of a particular type, such as those involving $f_{x_i x_j}^{(i,j)}$. A particular case of wide interest is the model which contains no interaction terms, often called a main-effect model,

$$E[y(x_1,\ldots,x_k)] = \mu + \sum_{i=1}^{k} f^{(i)}_{x_i}.$$  

The model used to describe an experiment with two factors, $F_1$ and $F_2$, with $F_1$ nesting $F_2$ is generally written as

$$E[y(x_1x_2)] = \mu + f^{(1)}_{x_1} + f^{(2)}_{x_1x_2}.$$  

In this case the number $f^{(2)}_{x_1x_2}$ is associated not only with the $x_2$-th level of factor $F_2$ but also with the $x_1$-th level of factor $F_1$. The nested factorial model for an experiment with $k$ factors $F_1,F_2,\ldots,F_k$ can be written

$$E[y(x_1,\ldots,x_k)] = \mu + f^{(1)}_{x_1} + f^{(2)}_{x_1x_2} + f^{(3)}_{x_1x_2x_3} + \ldots + f^{(k)}_{x_1\ldots x_k}.$$  

The models 2.14, 2.15 and 2.17 can be represented in the general form

$$E[y] = X\beta,$$

where $y$ is the vector of observations, $\beta$ is the vector of parameters, $\mu, f^{(i)}_{x_i}, f^{(2)}_{x_1x_j}$, and so on, and $X$ is the model matrix of zeros and ones.
A general feature of these models is that they are not of full rank. This means that no component of the vector $\beta$ is estimable. Linear classificatory models are generally made up of mixtures of models 2.14 and 2.17.

A. Orthogonality and Non-orthogonality

The simple case applicable to models of full rank is as follows,

\[ y = X\beta + \varepsilon = X_1^1 \beta_1 + X_2^2 \beta_2 + \varepsilon. \]

In this case we say the data-model situation is **orthogonal** with respect to the groups of parameters $\beta_1$ and $\beta_2$ if

\[ X_1^1 X_2 = \phi, \]

where $\phi$ denotes the null matrix.

This notion can be extended to models not of full rank. Suppose, for example, that

\[ y = X_1^1 \beta_1 + X_2^2 \beta_2 + \varepsilon \]

can be written in a full rank reparameterized form as

\[ y = W_1 \theta_1 + W_2 \theta_2 + W_3 \theta_3 + \varepsilon \]

where

\[ \theta_1 = A_1^1 \beta_1, \quad \theta_2 = A_2^2 \beta_2 \]

and $C(A_1^1)$ is a basis for estimable functions of $\beta_1$, $C(A_2^2)$ is a basis for estimable functions of $\beta_2$ and $W_1^1 W_2 = \phi$ with $W_3$ such that

\[ C(X_1^1, X_2^2) = C(W_1, W_2, W_3). \]
The data-model situation is then said to be orthogonal with respect to the parameters $\beta_1$ and $\beta_2$. It is worthwhile to note that the model 2.19 can always be reparameterized to

\[(2.22) \quad y = Z_1 \delta_1 + Z_2 \delta_2 + \epsilon\]

such that $Z_1'Z_2 = \phi$. However $Z_1$ is not, in general, a matrix function of $X_1$ alone and $Z_2$ is not, in general, a matrix function of $X_2$ alone.

Expression 2.22 is known as an orthogonal parameterization of the linear model 2.19.

Two similar concepts are often used to describe linear models. These are i) full-rank models and ii) maximal-rank models. Both of these concepts are closely related to the rank of the $n \times p$ model matrix $X$. A model is of full rank if the rank of $X$ is equal to $p$. A multifactorial data-model situation is of maximal rank if the rank of the model matrix is equal to the rank of the model matrix that would occur if there were an observation for every cell in the associated complete multifactorial partition.

A multifactorial data-model situation is of submaximal rank if the rank of the model matrix is less than the rank of the model matrix that would occur if there were an observation in every cell of the associated complete multifactorial partition.

We will now examine the property of non-orthogonality with specific reference to models of full rank, maximal rank and submaximal rank. Since a model of full rank is also of maximal rank, we will use the terms maximal and submaximal to apply only to models not of full rank.

The product of two arbitrary matrices $A'B$ may be non-null for three basically different reasons. Let $O_0$ represent a basis for $C(A) \cap C(B)$. Let $O_1$ be its extension to a basis for $C(A)$ and let $O_2$ be an extension
of $0_0$ such that $(0_0, 0_2)$ forms a basis for $C[B]$. Then the product $A'B$ will be non-null whenever

\begin{align*}
\text{i) } & 0_0 = \phi \text{ and } C[0_1] \text{ is not orthogonal to } C[0_2], \\
\text{ii) } & 0_0 \neq \phi \text{ and } C[0_1] \text{ is not orthogonal to } C[0_2], \\
\text{iii) } & 0_0 \neq \phi \text{ and } C[0_1] \text{ is orthogonal to } C[0_2].
\end{align*}

Any data-model situation that is non-orthogonal with respect to the parameters $\beta_1$ and $\beta_2$ can be characterized by one of the above situations. The first property is generally associated with models of maximal rank and the last two with models of submaximal rank. There are exceptions, however, to both of these statements.

The reasons for introducing conditions 2.23 will be clarified in Chapter IV. At this point we will only say that inferences drawn from these three situations are basically different and should be treated as such. In order to find solutions for models of submaximal rank, it is often necessary to impose conditions that are not standard nor easy to determine.

## B. Models of Maximal Rank

The works of Potthoff (1962a, 1962b) and Bradu (1965) apply to maximal rank, additive factorial models. Potthoff (1962b) assumes the following model,

\[ E[y] = X_1 t_1 + X_2 t_2 + X_3 t_3 + X_4 t_4, \]

where $y$ is an $n \times 1$ vector of observations, $X_i$ is an $n \times p_i$ design matrix, and $t_i$ is a $p_i \times 1$ vector of treatment parameters, $i = 1, 2, 3, 4$. The development in Potthoff (1962a) is essentially the same. The following definitions are made.
1) $H_{ij} = X_i^j X_j^i$, $i,j = 1,2,3,4$.

ii) $R_i = X_i^i y$, $i = 1,2,3,4$.

iii) $Q_1$ is a vector of the form

\begin{equation}
Q_1 = R_1 + A_{12} R_2 + A_{13} R_3 + A_{14} R_4 ,
\end{equation}

where $A_{ii}$ is $p_1 x p_1$, $i = 2,3,4$, with

\begin{equation}
E[Q_1] = C_{11} t_1 ,
\end{equation}

and $C_{11}$ is $p_1 x p_1$ of rank $p_1 - 1$. That is, $E[Q_1]$ is a vector of estimable functions of $t_1$ alone. The normal equations can be written

\begin{equation*}
\begin{align*}
H_{11} t_1 + H_{12} t_2 + H_{13} t_3 + H_{14} t_4 &= R_1 \\
H_{21} t_1 + H_{22} t_2 + H_{23} t_3 + H_{24} t_4 &= R_2 \\
H_{31} t_1 + H_{32} t_2 + H_{33} t_3 + H_{34} t_4 &= R_3 \\
H_{41} t_1 + H_{42} t_2 + H_{43} t_3 + H_{44} t_4 &= R_4
\end{align*}
\end{equation*}

A degree of balance is also introduced by assuming that, for any factor, each level appears an equal number of times in the design. Potthoff then exhibits classes of matrices $H_{ij}$, $A_{ij}$, and $C_{ii}$ which produce tractable analyses, and for which matrices $X_i$ exist. For example, Design Class 1 for four-factor additive models is defined below.
Design Class 1

Let \( h = p_1 p_2 p_3 p_4 \), then

\[
\begin{align*}
H_{12} H_{23} &= (h^2/p_1 p_2 p_3) J_{13}, \\
H_{12} H_{24} &= (h^2/p_1 p_2 p_4) J_{14}, \\
H_{13} H_{34} &= (h^2/p_1 p_3 p_4) J_{14}, \\
H_{13} H_{32} &= (h^2/p_1 p_2 p_3) J_{12}, \\
H_{14} H_{42} &= (h^2/p_1 p_2 p_4) J_{12}, \\
H_{14} H_{43} &= (h^2/p_1 p_3 p_4) J_{13},
\end{align*}
\]

where \( J_{i,j} \) is the \( p_i \times p_j \) matrix of one's. For any design belonging to this class, if

\[
\begin{align*}
A_{12} &= -(p_2/h) H_{12} + (2/p_1) J_{12}, \\
A_{13} &= -(p_3/h) H_{13}, \\
A_{14} &= -(p_4/h) H_{14},
\end{align*}
\]

and

\[
C_{11} = (h/p_1) I_{p_1} + A_{12} H_{21} + A_{13} H_{31} + A_{14} H_{41}.
\]

Potthoff also discusses methods for calculating sums of squares appropriate for testing estimable functions of \( t_i \).

Bradu (1965) discusses much the same problem as Potthoff but in a more general sense. The model assumed is
where \( p \) is any positive integer. A method is derived similar to the Gram-Schmidt orthonormalization procedure that involves inverting matrices of order \( p_i \) instead of \( \sum p_i \). The conditions imposed on the parameters \( t_i \) to obtain unique solutions are

\[
(2.29) \quad \mathbf{1}'X_i t_i = 0 ,
\]

where \( \mathbf{1}_n \) is an \( n \times 1 \) vector of ones. The procedure is recursive and the matrices \( X_i \) and parameters \( t_i \) are replaced at each iteration by matrices \( X_i^{(k)} \) and estimates \( \hat{t}_i^{(k)} \). Thus, after \( p \) iterations the estimates \( \hat{t}_i^{(1)} \) (the iterations are counted backwards so that the results of the last iterations are \( X_i^{(1)} \) and \( \hat{t}_i^{(1)} \)) form a set of least squares solutions to the normal equations. Bradu also discusses the calculation of sums of squares and tests of hypotheses.

C. Models of Submaximal Rank

Submaximal rank models appear primarily in one area of the statistical literature, the area known as fractional replication. A fractional replicate is a subset of possible treatment combinations that one can obtain from a factorial experiment. A thorough discussion of this can be found in Kempthorne (1947, 1952). Fractional replicates are used in experiments where it is reasonable to assume that certain preassigned contrasts in the parameters are negligible. The unique feature of these designs is that...
some functions of main effects alone and interactions alone are not estimable, but their sum is estimable. This property is known as confounding. Consider, for example, a simple experiment with two factors $F_1$ and $F_2$ each at two levels, and suppose that the following treatment combinations are observed

<table>
<thead>
<tr>
<th>$F_1$</th>
<th>$F_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>level 1</td>
<td>level 1</td>
</tr>
<tr>
<td>level 2</td>
<td>level 2</td>
</tr>
</tbody>
</table>

where $x$ denotes an observation. Let the expectations be written as

$$E[y_{ij}] = \mu + f_{i1} + f_{j1} + f_{ij}^{(1,2)} \text{ with } \sum_i f_{i1}^{(1,2)} = \sum_j f_{ij}^{(1,2)} = 0$$

for $(i,j) = (1,1), (1,2), (2,2)$. Specifically, the model is

$$E[y_{11}] = \mu + f_{11} + f_{11}^{(1,2)}$$
$$E[y_{12}] = \mu + f_{12} + f_{12}^{(1,2)}$$
$$E[y_{22}] = \mu + f_{22} + f_{22}^{(1,2)}$$

It is relatively easy to see that no linear combination of these observations has expectation
Using the conditions $\sum_{i} f_{ij}^{(1,2)} = \sum_{j} f_{ij}^{(1,2)} = 0$, and calculating $E[y_{22} - y_{11}]$ gives

\[(2.30) \quad E[y_{22} - y_{11}] = f_{2}^{(1)} - f_{1}^{(1)} + f_{2}^{(2)} - f_{1}^{(2)} .\]

In other words, the main-effect contrast $f_{2}^{(1)} - f_{1}^{(1)}$ is confounded with the main-effect contrast $f_{2}^{(2)} - f_{1}^{(2)}$.

Fractional factorial designs are constructed so that main-effects are confounded with higher order interactions instead of other main-effects. Various papers exist describing useful fractional factorial designs for certain numbers of factors and factor levels. One such paper is that by Addelman and Kempthorne (1961).

The concept of confounding is one that can apply to other types of designs. One of the main problems associated with non-orthogonal models and missing observations is unplanned confounding that enters into the estimation process.

D. An Aspect of Reparameterization

In this section one aspect of the process known as reparameterization is examined. It is introduced here because for non-orthogonal models it illustrates a serious problem for the estimation process.

Consider the simple two-factor nested model, written in a more standard form than 2.17 as
where \( i = 1,2,...,t, \ j = 1,2,...,s, \ k = 1,2,...,r. \) It can be seen that any column of the model matrix corresponding to \( \alpha_i \) can be constructed from appropriate columns corresponding to the \( \beta_{ij}. \) It is common practice to impose conditions on \( \beta_{ij} \) such as

\[
(2.32) \quad \sum_j \beta_{ij} = 0
\]

for all \( i \) in order to obtain unique solutions to the normal equations. New parameters satisfying 2.32 are then estimated. With this in mind, consider the linear function of the observations

\[
(2.33) \quad a'y = y_i.. - y_m..
\]

where the dot (\( \cdot \)) signifies taking the mean over the corresponding subscript. Clearly \( a \) is contained in \( C(X); \) and therefore, \( a'y \) is the BLUE of its expectation,

\[
(2.34) \quad E[a'y] = \alpha_i - \alpha_m + \frac{1}{sr} \left( \sum_j \beta_{ij} - \sum_j \beta_{mj} \right).
\]

Because of conditions 2.32, it seems reasonable to state that \( a'y \) is the BLUE of \( \alpha_i - \alpha_m. \) When any other set of conditions is imposed, differences of treatment means no longer yield unbiased estimates of simple differences of treatment effects. In this way the conditions imposed on the parameters
β_{ij} appear in the estimation of linear functions of the parameters α_i. The fact that this occurs and needs to be stressed appears in Elston and Bush (1964). They make the distinction between what they term restrictions and side conditions. The model they discuss is a two-factor model with interaction. In that situation the main-effects nest certain interactions. Conditions imposed on the interaction parameters appear in the estimation of main-effects and are therefore termed restrictions. Other conditions used in obtaining unique solutions are termed side conditions. When some data are missing, the restrictions imposed to obtain unique solutions influence the estimation process and therefore determine tests of hypotheses that can or cannot be made. As will be demonstrated later, this concept is critical to the analysis of any unplanned non-orthogonal experiment, and seems to be all but ignored by other authors.

E. Computational Procedures

In the past few years several algorithms for analysing general non-orthogonal experimental designs have appeared. Three of these, Wilkinson (1970), Fowlkes (1969) and Bradley (1968) will be examined in some detail.

1. Wilkinson's algorithm

Because of the unique nature of this algorithm, more detail is presented than in those that follow. The procedure itself is recursive in nature. Subsets of the parameters are fitted to the model

\[
y = X\beta + \epsilon, \quad (2.35)
\]
where $y$ is an $n \times 1$ vector of observations, $X$ is an $n \times p$ design matrix, $\beta$ is a $p \times 1$ vector of parameters, and $\varepsilon$ is an $n \times 1$ vector of random variables distributed as $\text{MVN}(0, \sigma^2 I)$. The general approach is to partition the matrix $X$ into

\begin{equation}
X = (X_1, X_2, \ldots, X_q),
\end{equation}

where $X_i$ is $n \times p_i$ and then fit successively, adjusting previous fits so that at the $k$-th stage the solutions are a set of least squares solutions for the linear model incorporating $X_1, X_2, \ldots, X_k$.

The algorithm can be described briefly as follows. Calculate a least squares solution, $\hat{\beta}_1$, for the submodel, $E[y] = X_1 \beta_1$. Using the residual, $y - X_1 \hat{\beta}_1$, calculate the vector $\hat{\beta}_2$, a partial least squares solution to the model, $y = X_1 \beta_1 + X_2 \beta_2$. From $\hat{\beta}_1$ and $\hat{\beta}_2$, calculate the vector $\hat{\beta}_1$ such that the composite vector, $\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$, is a least squares solution to the model, $y = X_1 \beta_1 + X_2 \beta_2$. This process can be continued similarly to calculate $\hat{\beta}_3, \hat{\beta}_4$, etc. in a recursive manner.

A detailed derivation of the least squares basis for this algorithm follows. The basic model can be written as

\begin{equation}
y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon.
\end{equation}

Assume $\hat{\beta}_1$ is any solution to the normal equations,

\begin{equation}
X_1'X_1 \hat{\beta}_1 = X_1' y.
\end{equation}
The vector $\tilde{\beta}_1$ used by Wilkinson is

$$\tilde{\beta}_1 = C_1 X'_1 y,$$

where $C_1$ satisfies $(X'_1 X_1)C_1(X'_1 X_1) = X'_1 X_1$. The residuals can be written as

$$y - X_1 \tilde{\beta}_1 = y - X_1 C_1 X'_1 y = (I - M_1) y = R_1 y.$$

The normal equations for the full model 2.37 are

$$\begin{bmatrix} X'_1 X_1 & X'_1 X_2 \\ X'_2 X_1 & X'_2 X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} X'_1 y \\ X'_2 y \end{bmatrix}. $$

These give reduced normal equations for $\beta_2$ eliminating $\beta_1$ as

$$X'_2 X'_1 \beta_2 = X'_2 y \quad \text{or} \quad X'_2 X'_1 X_1 \beta_2 = X'_2 X_1 y.$$

Let $C_2$ be any conditional inverse of $X'_2 X'_1 X_1$, that is

$$\begin{bmatrix} X'_1 X_1 \\ X'_2 X_1 \\ X'_2 X_2 \end{bmatrix} C_2 \begin{bmatrix} X'_1 X_1 \\ X'_2 X_1 \\ X'_2 X_2 \end{bmatrix} = X'_2 X_1 X_2.$$

Then, a solution to equations 2.42 is

$$\hat{\beta}_2 = C_2 X'_2 y.$$
This gives fitted values for $R_1X_2\hat{\beta}_2$ as

\[(2.45) \quad R_1X_2\hat{\beta}_2 = R_1X_2\beta_2,\]

and the least squares predicted values for the full model 2.37 can be written as

\[(2.46) \quad \hat{y} = X_1\tilde{\beta}_1 + R_1X_2\hat{\beta}_2.\]

These predicted values are usually written as

\[(2.47) \quad \hat{y} = X_1\tilde{\beta}_1 + (I-M_1)X_2\hat{\beta}_2\]
\[= X_1\tilde{\beta}_1 - M_1X_2\hat{\beta}_2 + X_2\hat{\beta}_2\]
\[= X_1(\tilde{\beta}_1 - C_1X'X_2\hat{\beta}_2) + X_2\hat{\beta}_2.\]

A least squares solution for $\beta_1$ is $\hat{\beta}_1 = \tilde{\beta}_1 - C_1X'X_2\hat{\beta}_2$; therefore $\hat{y}$ can be written as

\[(2.48) \quad \hat{y} = X_1\hat{\beta}_1 + X_2\hat{\beta}_2.\]

The least squares residuals can be expressed as

\[(2.49) \quad y - \hat{y} = y - X_1\tilde{\beta}_1 - R_1X_2\hat{\beta}_2\]
\[= R_1y - R_1X_2\hat{\beta}_2\]
\[= R_1(y - R_1X_2\hat{\beta}_2).\]
Substituting expression 2.44 for $\hat{\beta}_2$ gives

\[
(2.50) \quad y - \hat{y} = R_1 (I - X_2 C_2 X'_2) R_1 y
\]

\[
= R_1 S_2 R_1 y,
\]

where $S_2 = [I - X_2 C_2 X'_2]$ is Wilkinson's sweep operator. Notice that $S_2$ is not in general idempotent. The recursive nature of the algorithm begins to become apparent at this point. The procedure is to obtain the residuals

\[
z_0 = R_1 y
\]

from fitting the first part of the model and then operate on $z_0$ with the matrix

\[
(2.51) \quad S_2 = I - X_2 C_2 X'_2.
\]

This gives

\[
(2.52) \quad S_2 z_0 = S_2 R_1 y = y - X_1 \hat{\beta}_1 - X_2 \hat{\beta}_2
\]

which is termed the vector of apparent residuals, which are in general not unique. Finally, operate on vector $S_2 R_1 y$ with the operator $R_1$. For most designs the second operation with $R_1$ can either be deleted or simplified when orthogonalities in the design are taken into account. The final residuals can be written as
(2.53) \[ R_2 y = R_1 s_2 R_1 y. \]

where \( R_2 \) is the residual operator for fitting \( X_1 \) and \( X_2 \). The process can be repeated using \( R_2 \) in place of \( R_1 \) and affixing \( X_3 \beta_3 \) to the model.

The next step in the development is to evaluate the sweep operator \( S_k \) for the \( k \)-th stage. That is, consider the linear model

(2.54) \[ y = (X_1 \ldots X_p) \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \epsilon. \]

Assume the matrices

\[ R_1, R_2, \ldots, R_{k-1} \]
\[ S_2, \ldots, S_{k-1} \]

are known where

(2.55) \[ R_i = R_{i-1} s_i R_{i-1} \]

and

\[ S_i = I - X_i (X_i' R_{i-1} X_i)^{-1} X_i' \]

for \( i \) such that \( 1 \leq i \leq k-1 < p \). The least squares residuals from fitting \( X_1, \ldots, X_{k-1} \) are

(2.56) \[ z_{k-2} = R_{k-1} y. \]
The object then is to construct $\mathbf{R}_k$, $\mathbf{S}_k$ and $\mathbf{\hat{B}}_k$, and to extend the fit to include $X_k \beta_k$. This is accomplished using the matrix

$$Q_k = R_{k-1} M_k R_{k-1}$$

where $M_k = X_k (X_k' X_k)^{-1} X_k'$. The matrix $Q_k$ is called the shrinkage operator by Wilkinson. To construct $R_k$, $S_k$ and $\beta_k$ use is made of the reduced minimal polynomial of the matrix $Q_k$. The reduced minimal polynomial $p_k(x)$ is defined to be the polynomial of minimal degree with the constant term normalized to one such that

$$q_k(p_k(Q_k)) = \phi.$$ 

The nonzero roots of the polynomial $p_k(x)$ can be written as

$$\left( e^{-1}, e^{-1}, \ldots, e^{-1} \right).$$

The vector

$$\left( e_1, e_2, \ldots, e_K \right)$$

is the vector of distinct nonzero characteristic roots of the matrix $Q_k$, and the elements are called *efficiency factors* by Wilkinson. The reason that emphasis is placed on the reduced minimal polynomial of the matrix $Q_k$ is demonstrated in the following theorem. (The proof is given here since it is not in the paper.)
Theorem 2.1 Given the notation above, define

\[ q_k(x) = \frac{1 - p_k(x)}{x}. \]

Let \( A_k = X'_k R_{k-1} X_k \) and \( C_k^- \) be a conditional inverse of \( X'_k X_k \). Then the matrix

\[ \tilde{C}_k = C_k^- q_k(A_k C_k^-) \]

is a conditional inverse of \( A_k \).

Proof The theorem will be proved by verifying that

\[ A_k C_k^- A_k = A_k. \]

Let

\[ A_k C_k^- A_k = X'_k R_{k-1} X_k C_k^- q_k(X'_k R_{k-1} X_k C_k^-)X'_k R_{k-1} X_k \]

\[ = X'_k R_{k-1} M_{R_{k-1}} q_k(R_{k-1} M_{R_{k-1}})X_k \]

\[ = X'_k R_{k-1} [I - p_k(Q_k)]X_k \]

\[ = X'_k R_{k-1} X_k - X'_k R_{k-1} p_k(Q_k)X_k. \]

Since \( M_{k} X_k = X_k \) and \( R_{k-1} p_k(Q_k) = R_{k-1} p_k(Q_k) R_{k-1} \), the second term in the preceding expression can be written as
The quantity \( R_{k-1} P_k(Q_k) R_k^{-1} X_k \) when multiplied by its transpose gives

\[
(2.64) \quad [R_{k-1} P_k(Q_k) R_k^{-1} M_k] [R_{k-1} P_k(Q_k) R_k^{-1} M_k]' = R_{k-1} P_k(Q_k) Q_k P_k(Q_k) R_{k-1} = \phi.
\]

Therefore it must follow that

\[
(2.65) \quad R_{k-1} P_k(Q_k) R_k^{-1} M_k = \phi
\]

and \( A_k \tilde{C}_k A_k = A_k \).

When applying this result there are basically two situations that can arise.

1. The reduced minimal polynomial \( p(x) \) is known.
2. The reduced minimal polynomial \( p(x) \) is not known.

If \( p(x) \) is known from previous experience or can be derived from the combinatorics of the design, the problem is simply to evaluate the polynomial with the proper argument to obtain the estimates \( \hat{\beta}_k \) and the residuals which are in fact \( R_k y \).

In the second case, which will be the common one, Wilkinson describes an adaptive or learning mode of the algorithm which constructs the reduced minimal polynomial and determines the order of balance (the degree of the reduced minimal polynomial). There are several ways to formulate this
process depending upon the particular form chosen for expressing the polynomial \( p_k(x) \). The basic procedure is related to a set of procedures described in Faddeev and Faddeeva (1963), Chapter VI.

The first step is to construct a dummy variate \( \delta_k \). This can be done by calculating

\[
\delta_k = R_{k-1} X_k \psi_k
\]

where the vector \( \psi_k \) has elements \( \psi_{ki} = \xi^i \), for some transcendental number. Wilkinson states that a suitable rational number will work for a computer implementation. Let

\[
z_0 = \delta_k
\]

and define

\[
\gamma_1 = X_k^t z_0 = \tilde{\gamma}_1.
\]

Next calculate

\[
k^e_1 = \frac{\gamma_1 c_k \gamma_1}{\bar{z}_0 z_0},
\]

\[
t_1 = \frac{1}{\bar{c}_k} \gamma_1
\]

where

\[
c_k^{-1} = (X_k^t X_k)^{-1}
\]
and

(2.70) \[ z_1 = z_0 - R_{k-1}^* x_{k-1}^* \]
\[ = k_{1}^* z_{0} \]

The above procedures can be repeated giving

(2.71) \[ \gamma_2 = x_{k-1}^* z_1 \]

(2.72) \[ \tilde{\gamma}_2 = \gamma_2 - b_{21} y_1 \]

where

\[ b_{21} = (\gamma_2^* c_{k} y_1)/(\gamma_1^* c_{k} y_1) \]

(2.73) \[ \tilde{k}_{k}^2 = \frac{\tilde{\gamma}_2^* c_{k} y_2}{\tilde{z}_1^* z_1} \]

(2.74) \[ t_2 = \frac{1}{k_{k}^2} c_{k}^* y_2 \]

and

(2.75) \[ z_2 = k_{2}^* z_1 = z_1 - R_{k-1}^* x_{k-1}^* \]
We define recursively as follows:

(2.76) \[ \gamma_i = X_{k_i} z_{i-1} \]

(2.77) \[ \tilde{\gamma}_i = \gamma_i - \Sigma_{l=1}^{i-1} b_{i,l} \gamma_l \]

where

\[
\begin{pmatrix}
\gamma_1^{C_1 y_1} & \cdots & \gamma_1^{C_1 y_{i-1}} \\
\gamma_2^{C_1 y_1} & \cdots & \gamma_2^{C_1 y_{i-1}} \\
\vdots & \ddots & \vdots \\
\gamma_{i-1}^{C_1 y_1} & \cdots & \gamma_{i-1}^{C_1 y_{i-1}} \\
\end{pmatrix}
\begin{pmatrix}
b_{i1} \\
b_{i2} \\
\vdots \\
b_{i,i-1} \\
\end{pmatrix}
= 
\begin{pmatrix}
\gamma_1^{C_1 y_1} \\
\gamma_1^{C_1 y_2} \\
\vdots \\
\gamma_1^{C_1 y_{i-1}} \\
\end{pmatrix}
\]

(2.78) \[ k_i e_i = (\tilde{\gamma}_i^{C_1 y_i})/(z_{i-1} z_{i-1}) \]

(2.79) \[ t_i = \frac{1}{k_i e_i} c_i^{C_1 y_i} \]

(2.80) \[ z_i = z_{i-1} - R_{k_i} X_i t_i \]
Wilkinson states in his paper that the values of $b_{j\ell}$ are the elements of the matrix $C_k^-$. It is this author's feeling that the above must be a misprint, since there does not appear to be any justification for the statement as given. If it were the case, the values of $b_{j\ell}$ would clearly be independent of the particular dummy variate chosen, which contradicts another statement by Wilkinson. The above recursive procedure can be repeated until the vector $z_K = \phi$, then $K$ will be the order of balance. When the process is repeated using the vector $R_{k-1}Y$ in place of $\delta_k$, the final vector $z_k$ will be the vector of least squares residuals, and a least squares estimate for the vector $\beta_k$ will be

$$\hat{\beta}_k = \sum_{i=1}^{K} t_i .$$

It should be kept in mind that when the fit is extended to include $X_{k+1}$ the quantities

$$z_0, \ldots, z_K$$

$$\sim \sim$$

$$k^e_1, \ldots, k^e_K$$

$$t_1, \ldots, t_K$$
will each take on different values. However all previously calculated $\tilde{e}_i$ must be kept available to calculate quantities involving $R_{k-1}$.

The algorithm gives no information concerning degrees of freedom and for balanced experiments imposes conditions of the form $\sum a_i = 0$ or $\sum b_j = 0$. More will be said concerning both of these properties in later chapters, as well as what appears to be a slightly simpler form of the recursive procedure itself.

To aid in understanding some of the details of Wilkinson's algorithm the following simple incomplete experiment will be analysed using symbolic observations $y_{ij}$ instead of actual numeric values. Consider the following two-factor experiment from Mexas (1970). Observed treatment combinations are indicated with an $X$ in the following array.

<table>
<thead>
<tr>
<th>Factor $F_1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1^{(1)}$</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_1^{(2)}$</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>$f_2^{(1)}$</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_2^{(2)}$</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>$f_3^{(1)}$</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_3^{(2)}$</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$f_4^{(1)}$</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>$f_4^{(2)}$</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
The model for one replicate can be written in matrix form as

\[
\begin{pmatrix}
    y_{11} \\
    y_{13} \\
    y_{22} \\
    y_{24} \\
    y_{31} \\
    y_{33} \\
    y_{42} \\
    y_{44}
\end{pmatrix} = 
\begin{pmatrix}
    1 & 1 & 0 & 0 & 0 \\
    1 & 0 & 0 & 0 & 0 \\
    1 & 0 & 1 & 0 & 0 \\
    1 & 0 & 0 & 1 & 0 \\
    1 & 0 & 0 & 0 & 1 \\
    1 & 0 & 0 & 0 & 1 \\
    1 & 0 & 0 & 0 & 1 \\
    1 & 0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
    \mu \\
    f_1^{(1)} \\
    f_2^{(1)} \\
    f_3^{(1)} \\
    f_4^{(1)} \\
    f_1^{(2)} \\
    f_2^{(2)} \\
    f_3^{(2)} \\
    f_4^{(2)}
\end{pmatrix} + \varepsilon ,
\]

or equivalently,

\[
y = X_1 \mu + X_2 \beta_2 + X_3 \beta_3 + \varepsilon .
\]

The fit of \( y \) to \( X_1 \mu \) is quite simple and can easily be seen to give

\[
\tilde{\mu} = \frac{1}{8} \sum_{ij} y_{ij} = \ldots ,
\]

\[
R_1 = I_8 - \frac{1}{8} J_8 ,
\]

and
To determine the fit of $y$ to the model

(2.86) \[ E[y] = X_1 \mu + X_2 \beta + \varepsilon, \]

calculate the quantities

(2.87) \[ y_1 = X_2 z_0 \]

\[ = \begin{pmatrix}
  y_{11} - y_{..} + y_{13} - y_{..} \\
  y_{22} - y_{..} + y_{24} - y_{..} \\
  y_{31} - y_{..} + y_{33} - y_{..} \\
  y_{42} - y_{..} + y_{44} - y_{..}
\end{pmatrix}, \]
(2.88) \[ X_2^t X_2 = 2 I_4 \],

(2.89) \[ c_2^* = \frac{1}{2} I_4 \],

and

(2.90) \[ M_2 = X_2 c_2^* X_2^t = \frac{1}{2} X_2 X_2^t \].

Since the minimal polynomial of \( R M_2 R \) is unknown, the adaptive form of the analysis will be used. Since \( y_1 = \tilde{y}_1 \) the efficiency factor \( e_1 \) is

(2.91) \[ e_1 = \frac{y_1 c_2^* y_1}{z_1 z_1^t} \]

\[ = \frac{1}{2} \frac{[(y_{11} - y_{..}) + (y_{13} - y_{..})]^2 + \cdots + [(y_{14} - y_{..}) + (y_{h4} - y_{..})]^2}{(y_{11} - y_{..})^2 + (y_{13} - y_{..})^2 + \cdots + (y_{14} - y_{..})^2} \]

If the variate \( y \) represented a dummy variate as specified by (2.66) with \( R = 2 \), so that \( y \) is fitted exactly by \( X_2 \psi_2 \) for some \( \psi_2 \), then we would have

(2.92) \[ y_{11} - y_{..} = y_{13} - y_{..} \],

\[ y_{22} - y_{..} = y_{24} - y_{..} \],

\[ y_{31} - y_{..} = y_{33} - y_{..} \],

\[ y_{42} - y_{..} = y_{44} - y_{..} \].
Substituting these relationships in expression 2.91 gives

\[
(2.93) \quad e_1 = \frac{1}{2} \sum (y_{ij} - \bar{y})^2 = 1.
\]

It can be shown that the matrix \( R_{1} M_{2} R_{1} = \frac{1}{2} X_2 X_2^T - \frac{1}{6} J_6 \). For \( X_2 \) as in 2.81, the matrix \( R_{1} M_{2} R_{1} \) has unity as its only nonzero characteristic value, that value having multiplicity three. The current effects can be calculated as

\[
(2.94) \quad t_1 = e_1^{-1} X_2^T z_1 = \frac{1}{2} Y_1 = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ y_3 - \bar{y} \\ y_4 - \bar{y} \end{bmatrix},
\]

where \( y_i \) denotes the mean over the observations present. The first set of apparent residuals \( S_{1} z_1 \) can be calculated as

\[
(2.95) \quad S_{1} z_1 = z_1 - X_2^T t_1.
\]

It can be seen that if the vector \( y \) was in fact a dummy variate satisfying 2.92, the residuals 2.95 would be zero. This would indicate that the order of balance is unity and the reduced minimal polynomial of \( R_{1} M_{2} R_{1} \) is
Therefore, the effects in 2.94 are the estimates \( \hat{\beta}_1 \). Since the overall mean of the vector \( z_2 = s_1 z_0 \) is zero for all vectors \( y \), a "reanalysis step" applying \( R_1 \) would be redundant. From the form of 2.95 it can be seen that \( R_2 \) has the form

\[
R_2 = I - \frac{1}{2} (J_2 \otimes I_4),
\]

where \( J_2 \) denotes the \( 2 \times 2 \) matrix with all elements equal to unity and \( \otimes \) denotes the Kronecker product. In addition, the quantities

\[
X^T X_3 = 2 \, I_4
\]

\[
C_3 = \frac{1}{2} \, I_4
\]

can be evaluated, where \( C_3 \) denotes a conditional inverse (in this case the inverse) of \( X^T X_3 \). It then follows that \( R_2 M_3 R_2 \) has the value

\[
R_2 M_3 R_2 = \frac{1}{4} \, A \otimes B,
\]

where

\[
A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}
\]

and

\[
B = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}
\]
This implies that $R_2M_3R_2$ has characteristic values 1 and 0 with multiplicities 2 and 6, respectively. Therefore, the order of balance is 1 and $e_1$ is unity. The vector of current effects is $\hat{\beta}_2$ and has the value

\[ (2.101) \]

\[
\hat{\beta}_3 = \frac{1}{2} \begin{pmatrix}
 y_{1} - y_{3} \\
 y_{2} - y_{4} \\
 y_{3} - y_{1} \\
 y_{4} - y_{2}
\end{pmatrix}.
\]

The order of balance and $e_1$ could also have been determined using the properties of a dummy variate as was done for the vector $\beta_1$. From the forms of the operator $R_2$ and vector

\[ (2.102) \]

\[ z_3 = z_2 - x_3\hat{\beta}_3, \]

it can be shown that no adjustments of the vectors $\tilde{\mu}$ and $\hat{\beta}_2$ are necessary. Therefore, the vectors $\tilde{\mu}, \hat{\beta}_2$ and $\hat{\beta}_3$ are a set of least squares estimators. The final residuals are
(2.103) \[
\begin{pmatrix}
y_{11} - y_{13} + y_{33} - y_{31} \\
y_{13} - y_{11} + y_{31} - y_{33} \\
y_{22} - y_{24} + y_{44} - y_{42} \\
y_{24} - y_{22} + y_{42} - y_{44} \\
y_{31} - y_{33} + y_{13} - y_{11} \\
y_{33} - y_{31} + y_{11} - y_{13} \\
y_{42} - y_{44} + y_{24} - y_{22} \\
y_{44} - y_{42} + y_{22} - y_{24}
\end{pmatrix}
\]

It is clear that the elements of the vector \( \hat{\beta}_2 \) satisfy one linear constraint and that the elements of the vector \( \hat{\beta}_3 \) satisfy two linear constraints.

Perhaps the most serious drawback of this procedure is that it gives no information concerning degrees of freedom. The degrees of freedom can be determined from the multiplicities of the characteristic values introduced in the analysis. However, these multiplicities are not calculated by the algorithm. In addition, complex questions concerning estimable functions are ignored. More will be said concerning this algorithm and the above experiment in subsequent chapters.

2. Fowlkes' algorithm

The algorithm described by Fowlkes (1969) has essentially two aspects, model specification and computational procedure. The model is specified by means of the operational relations OR (+), CROSS(\text{*}) and NEST (+) among
the experimental factors. These relations permit a more compact model specification in some cases. For example, when the OR operator + is used, only those factors appearing are present in the analysis. When the CROSS operator * is used, the terms present in addition to all of their interactions appear in the analysis.

The computations are carried out by a series of operators CODE, ROW, MODEL, SCP, PATTER, DECOMP, SWP, and BAKSOL. The approach used to calculate the quantities involved in the analysis of variance is multiple linear regression. The above operators, designed to uncover linear relationships in singular matrices when they occur, are used in the computational steps of linear regression.

The operator CODE reparameterizes a vector corresponding to a level of one of the factors, either using $\sum a_i = 0$ or a relationship specified by the user. Other singularities are uncovered later in the process. The second operator ROW produces, from the operator ROW and the model, an actual row of the reparameterized model matrix. For example, ROW generates the interaction columns when interactions are specified in the model. The operator MODEL scans the model statement, validates the use of the operators +, * and +, and converts the model into a more usable form.

The remaining operators form the coefficient matrix of the normal equations and solve for the parameters. SCP forms the coefficient matrix of the normal equations, say $Z'Z$. The operator PATTER searches $Z'Z$ for block diagonal patterns and interchanges rows and columns if necessary. SWP performs one step in the Gauss elimination method for inverting a matrix. DECOMP performs an orthogonal decomposition of a matrix. For
example, the matrix $R$ is decomposed into

$$
\tilde{R} = (R_1, R_2)
$$

where $R_1$ is of full rank and has the same rank as $R$. In some cases $R_2$ may be null or not present. The last operator BAKSOL performs a back substitution in the triangular matrix generated by the Gauss elimination process to obtain the solutions. These operators also exist in versions appropriate for working on block diagonal matrices instead of entire matrices.

Aliased affects are indicated on the output by specifying the coefficients in the linear relationship. It is still not clear what is to be done with these relationships. If several effects are aliased the problem still remains of determining whether or not they are linearly independent.

3. Bradley's algorithm

Bradley (1968) describes an algorithm for finding a set of basis vectors for $C(X)$, where $X$ denotes the model matrix for the linear model

$$
y = X\beta + \varepsilon.
$$

Arbitrary data structures are classified into four broad classes depending upon the presence of missing cells and whether or not the model is a "full model," i.e. a full factorial model with all interactions present. The four classes are:
Class I. No empty cells; full model.
Class II. No empty cells; reduced model, i.e. some interaction parameters are not present.
Class III. Some empty cells; full model.
Class IV. Some empty cells; reduced model.

A set of rules is defined which partitions \( X \) (with perhaps some reordering) into

\[
X = (X_1, X_2),
\]

where \( X_1 \) is of full column rank and has the same rank as \( X \). For the most general class (Class IV), a matrix \( W^* \) is formed which generates a matrix \( W \). This in turn is used to construct a matrix \( Z \) from which \( X_1 \) can be generated. Symbolically,

\[
W^* \rightarrow W \rightarrow Z \rightarrow X_1.
\]

For all of the above classes one begins with only the distinct rows of the matrix \( X \). Let this matrix be called \( Z \). The matrix \( Z \) is considered because its rank is equal to the rank of \( X \). An example best illustrates this procedure. Consider the experimental arrangement below.

```
   b_1  b_2  b_3  b_4  b_5
a_1  X    X    X    
 a_2  
 a_3  X    
 a_4  X    X    
```
An X indicates that an observation was taken at a given level. The model will be a simple two way factorial with no interaction

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}. \]

The distinct rows of the matrix X are displayed below.

<table>
<thead>
<tr>
<th>ij</th>
<th>( \mu )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
<th>( \alpha_3 )</th>
<th>( \alpha_4 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( \beta_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first rule is to delete the column corresponding to \( \alpha_i \) for some \( i \) and \( \beta_j \) for some \( j \). This accomplishes one of the standard reparameterizations, and produces the matrix W below (where \( i = j = 1 \)).
Table 2.2. The matrix $W$.

<table>
<thead>
<tr>
<th>$ij$</th>
<th>$\mu$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For some models the above operations would be sufficient to generate an independent set of basis vectors for $\mathcal{C}(X)$. However, since singularities are still present the next rule is applied to $W$. For the most general class of experiments this rule produces the matrix $W^*$ as follows. Reorder the rows of $W$ so that the number of ones in each row does not decrease as one moves from top to bottom. For our example this gives $W^*$ as in Table 2.3.
Table 2.3. The matrix $W^*$.  

<table>
<thead>
<tr>
<th>$i$</th>
<th>$j$</th>
<th>$\mu$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The next rule which Bradley calls a "marking rule" is as follows. Mark (with an asterisk) the uppermost nonzero element in each column subject to the condition that no row can be marked more than once. Take in turn each column that has no marked element; eliminate elements by adding or subtracting marked columns until the uppermost element appears in an unmarked row, in which case it is marked, or until all elements in the column are zero. The initial marking gives the matrix below.

$$
\begin{pmatrix}
1^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1^* & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1^* & 0 \\
1 & 0 & 0 & 1^* & 0 & 0 & 0 & 0 \\
1 & 1^* & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1^* & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
\end{pmatrix}
$$
The "elimination" part of the marking rule gives the matrix

\[
Z_1 = \begin{bmatrix}
1^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1^* & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1^* & 0 \\
1 & 0 & 0 & 1^* & 0 & 0 & 0 & 0 \\
1 & 1^* & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1^* & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}
\]

Applying the same operations to the columns of the matrix \( X \) gives an independent set of basis vectors for \( C(X) \).

The most serious deficiency of this procedure is that the elimination process described by Bradley in the "marking rule" is not concise enough to be programmed in general on a computer. Another shortcoming is that after the elimination process the columns of the \( X \) matrix have been added and subtracted from one another and it is no longer clear what functions of the parameters are being estimated.
III. IMPOSING ESTIMABLE CONDITIONS

The reparameterizing process discussed in Chapter II involved imposing nonestimable conditions on the parameters. In the following chapters it will be necessary, from time to time, to impose estimable conditions on models (or submodels). However, this chapter really serves two purposes: first to derive some important results to be used later, and second to acquaint the reader with some basic procedures that are used throughout this dissertation. Some of these results appear also in Zyskind, Kempthorne et al. (1964).

A. Least Squares Derivation

Consider the linear model,

\( y = X\beta + \epsilon \),  \( (3.1) \)

where \( y \) is an \( n \times 1 \) vector of observations, \( X \) is an \( n \times p \) matrix of known constants, \( \beta \) is a \( p \times 1 \) vector of parameters to be estimated, and \( \epsilon \) is an \( n \times 1 \) vector of random variables with mean zero and variance \( \sigma^2 I \).

To obtain a least squares estimate for \( \beta \) subject to the estimable conditions

\( U\beta = c \),  \( (3.2) \)

where \( U \) is \( q \times p \) matrix of known constants and \( c \) is a \( q \times 1 \) vector of known constants, we use Lagrange multipliers and minimize
(3.3) \[ Q^* = \sum_i (y_i - \sum_j x_{ij} \beta_j)^2 + 2 \sum_k \rho_k (\sum_j x_{kj} \beta_j - c_k) \]
\[ = (y - X\beta)'(y - X\beta) + 2(UB - c)'\rho . \]

Taking derivatives and equating them to zero gives

(3.4) \[ \frac{1}{2} \frac{\partial Q^*}{\partial \beta_t} = -\sum_i (y_i - \sum_j x_{ij} \beta_j)x_{it} + \sum_k \rho_k U_{kt} = 0, \quad (t = 1, 2, \ldots, p). \]

These derivatives give the normal equations,

(3.5) \[ X'X\beta = X'y - U'\rho . \]

Since the conditions in Equation 3.2 are estimable, there exists a matrix $W$ such that

(3.6) \[ U = WX . \]

Therefore, we can write the normal equations 3.5 as

(3.7) \[ X'X\beta = X'(y - W'\rho) . \]

These are now "normal" type equations and hence, admit solutions. A set of solutions is

(3.8) \[ \hat{\beta} = (X'X)^{-1}X'(y - W'\rho) , \]
where \((X'X)^{-1}\) is any conditional inverse of \(X'X\), i.e. \(X'X(X'X)^{-1}X'X = X'X\).

To determine \(p\) such that \(UB = c\), we substitute \(\hat{\beta}\) in equation 3.2.

\[
\hat{UB} = U(X'X)^{-1}X'(y - W'p) = WX(X'X)^{-1}X'(y - W'p) \\
= WM_x(y - W'p) = c,
\]

where \(M_x\) denotes \(X(X'X)^{-1}X'\), the orthogonal projection operator on \(C(X)\).

Since the equations 3.2 are consistent we can write \(UB = c = WX\hat{\beta}_o\) for some vector \(\beta_o\). This gives equations 3.9 as

\[
WM_xW'p = WM_x(y - X\beta_o).
\]

The only condition for choosing the matrix \(W\) is that it satisfy \(U = WX\).

It will also be assumed that \(W = ZX'\) for some matrix \(Z\). Such a \(W\) always exists because \(U = WX\) can be written as \(U = WM_xX = WX(X'X)^{-1}X'X\).

Then a new matrix \(W\) can be chosen to be \(WX(X'X)^{-1}X'\). It also follows from the conjugate normal equations that such a \(W\) exists. Since \(WM_x = ZX'M_x = ZX' = W\), equations 3.10 can be written as

\[
WW'p = W(y - X\beta_o).
\]

These are also "normal" type equations and always admit solutions. One set of solutions is
Therefore, we can write the normal equations 3.7 as

\[(3.13) \quad X'X\hat{\beta} = X'[y - W'(WW')^{-1}W(y - X\hat{\beta}_0)]\]
or
\[(3.13) \quad X'X\hat{\beta} = X'[y - M_w(y - X\hat{\beta}_0)],\]

where $M_w$ is the orthogonal projection operator on $C(W')$. This implies that a set of solutions is

\[(3.14) \quad \hat{\beta} = (X'X)^{-1}X'[y - M_w(y - X\hat{\beta}_0)],\]

and the corresponding predicted values are

\[(3.15) \quad \hat{y} = X\hat{\beta} = M_x[y - M_w(y - X\hat{\beta}_0)].\]

Since $W = ZX'$ for some matrix $Z$, $M_w$ has the form $XZ'(ZX'XZ')^{-1}ZX'$. And since $MX = X$, it follows that $M_w^2 = MX = M_w$, which implies that the predicted values can be written as

\[(3.16) \quad \hat{y} = X\hat{\beta} = M_wy - M_w(y - X\hat{\beta}_0).\]

It should be mentioned that although the vector $\beta_o$ is not necessarily unique, the quantity $M_w\beta_o$ is unique. Therefore, the vector $y - X\beta_o$ may vary depending on the particular choice of $\beta_o$, but the vector $M_w(y - X\beta_o)$ will be independent of the choice of $\beta_o$. 
B. Degrees of Freedom, Sums of Squares, and Tests of Hypotheses

The minimum sum of squares can be written as:

\[(3.17) \quad Q^* = (y - X\hat{\beta})'(y - X\hat{\beta})\]
\[= (y - M_x y - M_w (y - X\beta_o))'(y - M_x y - M_w (y - X\beta_o))\]
\[= y'y - y'M_x y + (y - X\beta_o)' M_w (y - X\beta_o)\]
\[= y'(I - M_x) y + (y - X\beta_o)' M_w (y - X\beta_o).\]

If we assume that the vector of random variables \(\varepsilon\) mentioned previously is distributed as multivariate normal with mean zero and variance \(\sigma^2 I\), the two quadratic forms \(y'(I - M_x)y\) and \((y - X\beta_o)' M_w (y - X\beta_o)\) have chi-square distributions with the parameters listed below.

Table 3.1. Parameters of quadratic forms.

<table>
<thead>
<tr>
<th>Quadratic form</th>
<th>d.f.</th>
<th>Non-centrality parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y'(I - M_x)y)</td>
<td>rank ((I - M_x))</td>
<td>(\frac{1}{2} E(y'(I - M_x)E(y)))</td>
</tr>
<tr>
<td>((y - X\beta_o)' M_w (y - X\beta_o))</td>
<td>rank ((M_w))</td>
<td>(\frac{1}{2} E(y - X\beta_o)' M_w E(y - X\beta_o))</td>
</tr>
</tbody>
</table>

Since \((I - M_x)M_w = \phi\), the two quadratic forms are independent, and under the null hypothesis that \(y = X\beta\) and \(U\beta = \zeta\), both are central chi-squares.
C. The Variance of Linearly Estimable Functions

Suppose we wish to estimate \( \lambda' \beta \) where \( \lambda \) is a vector in the row space of \( X \), i.e. there exists a vector \( a \) such that \( \lambda = X'a \). The variance of \( \lambda' \beta \) can be written as

\[
\text{Var}[\lambda' \beta] = \text{Var}[\lambda' (X'X)^{-1} X' (y - M_y (y - X\beta_0))] .
\]

Since \( X\beta_0 \) is a constant,

\[
\text{Var}[\lambda' \hat{\beta}] = \text{Var}[\lambda' (X'X)^{-1} X' (I - M_y) y]
\]

\[
= \sigma^2 \lambda' (X'X)^{-1} X' (I - M_y) X (X'X)^{-1} \lambda
\]

\[
= \sigma^2 a' M_x (I - M_y) M_x a
\]

\[
= \sigma^2 a' (M_x - M_y) a .
\]

Clearly if \( a \) is a vector in the column space of \( W \), then \( \text{Var}[\lambda' \hat{\beta}] = 0 \). If \( a \) is a vector orthogonal to \( W \), then \( \text{Var}[\lambda' \hat{\beta}] = \sigma^2 a' M_x a \); and since \( a = X\gamma \) for some vector \( \gamma \), then

\[
\text{Var}[\lambda' \hat{\beta}] = \sigma^2 \gamma' X' X \gamma = \sigma^2 a'a .
\]

Since \( M_x X = X \), Equation 3.19 can be written as

\[
\text{Var}[\lambda' \hat{\beta}] = \sigma^2 a' [I - M_y] a .
\]
D. The Special Case When $c = 0$

For the special case when $c = 0$, which will be of interest later, the following relationships hold.

The normal equations are

\[(3.22) \quad X'X\hat{\beta} = X'(I - M_w)y,\]

and a set of solutions is

\[(3.23) \quad \hat{\beta} = (X'X)^{-1}X'(I - M_w)y.\]

The predicted values are

\[(3.24) \quad \hat{y} = X\hat{\beta} = (M_x - M_w)y,\]

which implies that the residuals are

\[(3.25) \quad y - \hat{y} = (I - M_x + M_w)y;\]

and the minimum sum of squares is

\[(3.26) \quad Q^* = y'y - y'M_x y + y'M_w y.\]

These results will be used to prove the following theorem.
Theorem 3.1 Let \( X_1 \) and \( X_2 \) be \( n \times p_1 \) and \( n \times p_2 \) matrices.

Define

\[
M_1 = X_1 (X_1' X_1)^{-1} X_1', \quad R_1 = I - M_1
\]

and

\[
M_2 = X_2 (X_2' X_2)^{-1} X_2', \quad R_2 = I - M_2.
\]

Assume also that there exists a matrix \( Z \) such that \( X_2 = X_1 Z \). Then a solution \( \hat{\beta} \) to the set of equations

\[
(3.27) \quad X_1^R_2 X_1 \hat{\beta} = X_1^R_2 y
\]

can be found by imposing the estimable conditions

\[
(3.28) \quad X_2' X_1 \hat{\beta} = \phi
\]

on the model

\[
(3.29) \quad y = X_1 \hat{\beta} + \varepsilon.
\]

Proof It will be sufficient to show that a set of solutions \( \tilde{\beta} \) obtained by imposing the estimable conditions 3.28 on the model 3.29 satisfies the equations 3.27. Let \( \tilde{\beta} \) be of the form

\[
(3.30) \quad \tilde{\beta} = (X_1' X_1)^{-1} X_1^R_2 y.
\]
From expression 3.23 it follows that $\tilde{\beta}$ is a least squares solution to model 3.29 and also satisfies conditions 3.28. The left-hand side of equation 3.27 becomes

$$\begin{align*}
X_1'^R_2X_1\tilde{\beta} &= X_1'(I - M_2)X_1(X_1'X_1)^{-1}X_1'(I - M_2)y \\
&= X_1'(I - M_2)M_1(I - M_2)y.
\end{align*}$$

Since $M_2M_1 = M_1M_2$, it follows that

$$\begin{align*}
(I - M_2)M_1 = M_1(I - M_2).
\end{align*}$$

Thus, 3.31 can be written as

$$\begin{align*}
X_1'^R_2X_1\tilde{\beta} &= X_1'M_1(I - M_2)(I - M_2)y.
\end{align*}$$

Because $R_2$ is idempotent and $X_1'M_1 = X_1$, then

$$\begin{align*}
X_1'^R_2X_1\tilde{\beta} &= X_1'R_2y.
\end{align*}$$

The equivalence established by this theorem will be a useful tool in later material.
Consider the linear model \( y = X\beta + \varepsilon \) written in the following form

\[
(4.1) \quad y = \sum_{i=1}^{p} X_i \beta_i + \varepsilon ,
\]

where each matrix \( X_i \) is \( n \times p_i \) and each vector of parameters has length \( p_i \). Alternatively, we can write

\[
(4.2) \quad y = (X_1 \ldots X_p) \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon .
\]

A linear model written in the above form will be referred to as a **partitioned linear model**. The normal equations which must be solved to obtain least squares estimates for the parameters have the following form

\[
(4.3) \quad \begin{bmatrix} X_{11}'X_{11} & X_{11}'X_{12} \ldots X_{11}'X_{1p} \\ X_{12}'X_{11} & X_{12}'X_{12} \ldots X_{12}'X_{1p} \\ \vdots \\ X_{p1}'X_{11} & X_{p1}'X_{12} \ldots X_{p1}'X_{1p} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} X_{11}'y \\ X_{12}'y \\ \vdots \\ X_{p1}'y \end{bmatrix} .
\]

This type of model forms the basis for the analysis of linear classificatory models.
In this chapter we examine in detail the estimation and analysis of a partitioned model with two partitions. No assumptions are made concerning the rank of the model or relationships between \( X_1 \) and \( X_2 \). A decomposition theorem is established which is useful for examining estimable functions and tests of hypothesis. The notation and some of the preliminary derivations closely follow that of Zyskind, Kempthorne, et al. (1964).

A. The Linear Model With Two Partitions

In the case \( p = 2 \), the partitioned linear model is

\[
y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon.
\]

The corresponding normal equations are

\[
\begin{align*}
\begin{bmatrix}
X'_1X_1 & X'_1X_2 \\
X'_2X_1 & X'_2X_2
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
&= 
\begin{bmatrix}
X'_1y \\
X'_2y
\end{bmatrix}.
\end{align*}
\]

To solve these equations we first write them as the two equations

\[
\begin{align*}
X'_1X_1 \beta_1 + X'_1X_2 \beta_2 &= X'_1y \\
X'_2X_1 \beta_1 + X'_2X_2 \beta_2 &= X'_2y.
\end{align*}
\]

Multiplying the first equation by \(-X'_21 (X'_1X_1)^{-1}\) and adding the result to the second equation gives the reduced normal equations for \( \beta_2 \), eliminating \( \beta_1 \).
\[(4.8) \quad x_2'[I - x_1(x_1'x_1)^{-1}x_1]x_2\beta_2 = x_2'[I - x_1(x_1'x_1)^{-1}x_1]y \]

And letting \( R_1 = I - x_1(x_1'x_1)^{-1}x_1 \), we have

\[(4.9) \quad x_2'R_1x_2\beta_2 = x_2'R_1y. \]

Since \( R_1 \) is symmetric and idempotent, these are "normal" type equations and always admit solutions. Let \( \hat{\beta}_2 \) be a solution to equations 4.9, then from equation 4.6 we have

\[(4.10) \quad \hat{\beta}_1 = (x_1'x_1)^{-1}x_1'(y - x_2\hat{\beta}_2). \]

The composite vector \( \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \) forms a solution to equation 4.5.

**B. Estimability**

For a model matrix \( X \) consisting of only one partition, a linear function of the parameters \( \lambda'\beta \) is estimable if and only if \( \lambda' \) belongs to the row space of \( X \). For a model with two partitions, the linear function

\[(4.11) \quad \begin{bmatrix} \lambda_1' \\ \lambda_2' \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \]

is estimable if and only if \( \begin{bmatrix} \lambda_1' \\ \lambda_2' \end{bmatrix} \) belongs to the row space of \( (X_1, X_2) \), or equivalently, the column space of
To determine linearly estimable functions we need to know the forms of $\lambda_1$ and $\lambda_2$ for which the following equations, called conjugate normal equations, are consistent,

\begin{equation}
\begin{pmatrix}
x_1'x_1 & x_1'x_2 \\
x_2'x_1 & x_2'x_2
\end{pmatrix}
\begin{pmatrix}
\theta_1 \\
\theta_2
\end{pmatrix}
= \begin{pmatrix}
\lambda_1 \\
\lambda_2
\end{pmatrix}.
\end{equation}

The first case to examine is that of $\lambda_2 = 0$. That is, what linear functions of $\beta_1$ alone are estimable? Solving for $\theta_2$ in terms of $\theta_1$ gives

\begin{equation}
\theta_2 = -(x_1'x_2)'x_1'x_1\theta_1.
\end{equation}

Substituting this result in the first set of equations in 4.13 gives

\begin{equation}
x_1'x_1\theta_1 + x_1'x_2(x_2'x_2)'x_2'x_1\theta_1 = \lambda_1.
\end{equation}

Letting $x_2(x_2'x_2)'x_2 = M_2$ and $(I - M_2) = R_2$, 4.15 becomes

\begin{equation}
x_1'R_2x_1\theta_1 = \lambda_1.
\end{equation}
These equations are consistent and admit solutions if and only if $\lambda_1$ is a vector in the column space of $X_1'X_1$, or equivalently, in the row space of $R_1X_1$. Using the same type of argument, a linear function $\lambda_2'\beta_2$ of the parameters $\beta_2$ alone is estimable if and only if $\lambda_2$ is a vector in the row space of $R_1X_2$.

We have thus far established necessary and sufficient conditions for the estimability of linear functions of each set of parameters $\beta_1$ or $\beta_2$ alone. This can be stated in the following theorem.

**Theorem 4.1**  Given the partitioned linear model

$$y = X_1\beta_1 + X_2\beta_2 + e,$$

a linear function of the parameters of the form

$$\lambda'\beta = (\lambda_1', 0)$$

is estimable if and only if $\lambda_1'$ is a linear combination of the rows of $R_2X_1$. Similarly, a linear function of the parameters of the form

$$\lambda'\beta = (0, \lambda_2')$$

is estimable if and only if $\lambda_2'$ is a linear combination of the rows of $R_1X_2$. 

Since \( C(X_1'R_2X_1) = R(X_2X_1) \), it follows that if the function

\[
(4.17) \quad \begin{bmatrix}
\beta_1 \\
(\lambda_1', \phi') \\
\beta_2
\end{bmatrix}
\]

is estimable, then

\[
(4.18) \quad \lambda_1 = X_1'R_2X_1\rho_1
\]

for some vector \( \rho_1 \). Equations 4.18 will be called the reduced conjugate normal equations for \( \beta_1 \) eliminating \( \beta_2 \). In other words the vector \( \lambda_1 \) can be written

\[
(4.19) \quad \lambda_1 = X_1'\eta_1 ,
\]

where \( \eta_1 \in C(R_2X_1) \). Similarly, if the function

\[
(4.20) \quad \begin{bmatrix}
\beta_1 \\
(\phi, \lambda_2') \\
\beta_2
\end{bmatrix}
\]

is estimable, then

\[
(4.21) \quad \lambda_2 = X_2'\eta_2 ,
\]

where \( \eta_2 \in C(R_1X_2) \). It is well known that if the linear function of the observations \( a'y \) is BLUE for \( \lambda' \beta \), then there is some \( \rho \) such that
(4.22) \[ \lambda = X'Xp = X'a. \]

Since \( R_2X_2 = \phi \), the conjugate normal equations for a linear function of \( \beta_1 \) alone can be written

\[
\begin{pmatrix}
X'_1X_1 & X'_1X_2 \\
X'_2X_1 & X'_2X_2
\end{pmatrix}
\begin{pmatrix}
\theta_1 \\
\theta_2
\end{pmatrix}
= \begin{pmatrix}
\lambda_1 \\
\phi
\end{pmatrix} = \begin{pmatrix}
X'_1 \\
X'_2
\end{pmatrix} R_2X_1 \rho_1. \tag{4.23}
\]

From this it follows that any linear function of the observations \( a'y \) with \( a \in C(R_1X_1) \) is BLUE for an estimable function of \( \beta_1 \) alone.

Similarly, any linear function of the observations \( b'y \) with \( b \in C(R_1X_2) \) is BLUE for an estimable function of \( \beta_2 \) alone.

Thus we have developed characterizations of BLUE's of estimable functions of \( \beta_1 \) alone and BLUE's of estimable functions of \( \beta_2 \) alone. The next question would seem to be, "Can there be a linear function of the observations, \( a'y \), that is BLUE for an estimable function \( \lambda_1\beta_1 + \lambda_2\beta_2 \), with \( a \notin C(R_1X_2) \) and \( a \notin C(R_2X_1) \)?" The answer to this question can be established by examining the conjugate normal equation 4.13. These equations are consistent for any \( \lambda_1 \) and \( \lambda_2 \) where \( \lambda_1 = X'_1\rho \) and \( \lambda_2 = X'_2\rho \).

Assume \( \rho \) satisfies the following

(4.24) \[ \rho = X_1\eta_1 = X_2\eta_2 \]

for some vectors \( \eta_1 \) and \( \eta_2 \). Clearly \( \rho \in C(X_1,X_2) \), and therefore \( \rho'y \) is BLUE for its expectation. Since \( \rho'R_1X_2 = \rho'R_2X_1 = \phi \), we have \( \rho \notin C(R_1X_2) \).
\( \rho \in \mathbb{C}(\mathbb{R}X_1) \). Therefore, a linear function of the observations \( \rho' y \), with \( \rho \in \mathbb{C}(\mathbb{R}X_1) \) and \( \rho \notin \mathbb{C}(\mathbb{R}X_2) \), exists that is BLUE for an estimable function \( \eta_1 \beta_1 + \eta_2 \beta_2 \) whenever \( \mathbb{C}(X_1) \cap \mathbb{C}(X_2) \neq \emptyset \).

It can be shown in the following theorem that the three spaces, \( \mathbb{C}(\mathbb{R}X_1), \mathbb{C}(\mathbb{R}X_2), \) and \( \mathbb{C}(X_1) \cap \mathbb{C}(X_2) \) characterize the BLUE's of all estimable functions.

**Theorem 4.2** Any nonzero vector \( a \) contained in \( \mathbb{C}(X_1, X_2) \), i.e. \( a = X_1 \eta_1 + X_2 \eta_2 \), has the unique decomposition

\[
(4.25) \quad a = R_2 X_1 \rho_1 + R_1 X_2 \rho_2 + z,
\]

where \( z \in \mathbb{C}(X_1) \cap \mathbb{C}(X_2) \), i.e. \( z = X_1 \theta_1 = X_2 \theta_2 \). In other words any linear function of the observations \( \rho' y \) that is BLUE for its expectation can be decomposed uniquely into the sum of a function that is BLUE for an estimable function of \( \beta_1 \) alone, a function that is BLUE for an estimable function of \( \beta_2 \) alone, and a function that is BLUE for an estimable function in both \( \beta_1 \) and \( \beta_2 \) that cannot be further decomposed.

**Proof** To establish the existence of such a decomposition it will be shown that

i) The three vector spaces \( \mathbb{C}(\mathbb{R}X_1), \mathbb{C}(\mathbb{R}X_2), \) and the intersection of \( \mathbb{C}(X_1) \) and \( \mathbb{C}(X_2) \) are disjoint.

ii) The span of these three spaces is a subset of \( \mathbb{C}(X_1, X_2) \).

iii) The sum of the dimensions of the three spaces is equal to the dimension of \( \mathbb{C}(X_1, X_2) \).

For notational convenience let the \( n \times n \) matrix \( M \) denote the orthogonal projection operator on the intersection of \( \mathbb{C}(X_1) \) and \( \mathbb{C}(X_2) \).
i) Since any vector in \( C(M) \) can be written as \( X_1 \eta_1 = X_2 \eta_2 \) for some \( \eta_1 \) and \( \eta_2 \), it is clear that \( M \) is orthogonal to both \( R_{\perp}X_1 \) and \( R_{\perp}X_2 \). Suppose that there exists a nonzero vector \( a \) in both spaces \( C(R_{\perp}X_1) \) and \( C(R_{\perp}X_2) \), then

\[
a'a = \eta_1^t X_{\perp}^t R_{\perp} X_{\perp} \eta_1 = \eta_2^t X_{\perp}^t R_{\perp} X_{\perp} \eta_1 = 0.
\]

Therefore, \( a \) must be the zero vector; and the three spaces, \( C(R_{\perp}X_1) \), \( C(R_{\perp}X_2) \), and \( C(M) \), must be disjoint.

ii) It can be seen from the definition of \( R_{\perp}X_1 \), \( R_{\perp}X_2 \) and \( M \) that any vector made up of a linear combination of the columns of these three matrices has the form \( X_1 \eta_1 + X_2 \eta_2 \). Therefore, the space of all such linear combinations must form a subspace of \( C(X_1, X_2) \).

iii) Let \( q \) be the dimension of \( C(X_1, X_2) \), denoted by \( \dim[C(X_1, X_2)] \), similarly define \( q_1 = \dim[C(X_1)] \), \( q_2 = \dim[C(X_2)] \), and \( q_{12} = \dim[C(M)] \). Let the columns of the matrix \( O \) be linearly independent and form a basis for \( C(M) \). Let \( O_1 \) be an extension of this basis so that \( (O_1, O_0) \) forms a basis for \( C(X_1) \). Clearly \( O_0 \) has \( q_{12} \) columns and \( O_1 \) has \( q_1 - q_{12} \) columns. Now consider \( R_2(O_1, O_0) = R_2(O_1, \phi) \). Since \( (O_1, O_0) \) forms a basis for \( C(X_1) \), there exist matrices \( B \) and \( D \) such that \( (O_1, O_0) = X_1 B \) and \( (O_1, O_0)D = X_1 \). Therefore, any vector in \( C(R_2X_1) \) can be written as \( R_2X_1 \eta_1 = R_2(O_1, O_0)D \eta_1 \) for some vector \( \eta_1 \). It follows from the definition of matrix \( B \) that any vector in \( C(R_2(O_1, O_0)) \) can also be written as \( R_2X_1 \rho_1 \) for some vector \( \rho_1 \). Hence, the matrix \( R_2(O_1, O_0) \) forms a basis for \( C(R_2X_1) \).
It can be shown by contradiction that the \( q_1 - q_{12} \) columns of \( R_1 O_1 \) are linearly independent. Assume they are not linearly independent; then there exists a nonzero vector \( \eta \) such that

\[
R_2 O_1 \eta = \phi.
\]

This implies that

\[
O_1 \eta = M_2 O_1 \eta.
\]

Since the columns of \( O_1 \) are linearly independent, the vector \( O_1 \eta \) is nonzero and is contained in the column space of \( X_2 \). This implies that \( O_1 \eta \) is in both \( C(X_1) \) and \( C(X_2) \) which is impossible because of the construction of \( O_1 \). Therefore, we have a basis for \( C(R_2 X_1) \) consisting of \( q_1 - q_{12} \) linearly independent columns, and \( \dim[C(R_2 X_1)] = q_1 - q_{12} \).

By a similar argument it can be shown that \( \dim[C(R_1 X_2)] = q_2 - q_{12} \).

Therefore,

\[
\dim[C(R_2 X_1)] + \dim[C(R_1 X_2)] + \dim[C(M)]
\]

\[
= q_1 - q_{12} + q_2 - q_{12} + q_{12} = q_1 + q_2 - q_{12}
\]

\[
= \dim[C(X_1, X_2)].
\]

The fact that the decomposition is unique for any estimable function \( \lambda_1 \beta_1 + \lambda_2 \beta_2 \) follows from the fact that the three spaces are disjoint and the original vector \( a \) is unique.
It is worth noting that the above decomposition is only a partitioning of the estimation space. In order to specify the total decomposition of any \( n \times 1 \) vector \( \alpha \), the error space must also be included. That is, any \( n \times 1 \) vector \( \alpha \) can be written as

\[
\alpha = \alpha_1 + \alpha_2,
\]

where \( \alpha_1 \in C(X_1, X_2) \) and \( \alpha_2 \in C_1(X_1, X_2) \), then \( \alpha_1 \) can be further decomposed as in Theorem 4.2.

Another description of the decomposition can be given in terms of null spaces. The null space of a matrix \( A \) is defined to be the set of all vectors \( x \) such that \( Ax = \phi \). Clearly, from the definition of the orthogonal complement, the \( N(A) \) is the same as \( C_1(A') \). Therefore the error space is \( N\begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \). The space of all vectors \( x \) such that \( x'y \) is BLUE for an estimable function in \( \beta_1 \) alone is \( N(X_2') \cap C(X_1, X_2) \). Similarly the space of BLUE's of estimable functions of \( \beta_2 \) is \( N(X_1') \cap C(X_1, X_2) \).

The four spaces in the total decomposition are:

\[(4.26) \quad N(X_2') \cap C(X_1, X_2): \text{BLUE's of estimable functions of } \beta_1 \text{ alone},\]

\( N(X_1') \cap C(X_1, X_2): \text{BLUE's of estimable functions of } \beta_2 \text{ alone},\)

\( C(X_1) \cap C(X_2): \text{BLUE's of estimable functions of both } \beta_1 \text{ and } \beta_2, \text{ and} \)

\( N\begin{bmatrix} X_1' \\ X_2' \end{bmatrix}: \text{error space.} \)
Theorem 4.2 establishes a decomposition of the column space of $X$ into disjoint and sometimes orthogonal subspaces. Each subspace represents a set of BLUE's for estimable functions of the parameters. The corresponding decomposition of the space of vectors $\lambda$ such that $\lambda'\beta$ is estimable can be similarly decomposed by examining the following expectations:

$$
E[X_1^Ry] = X_1^R\beta_1 + \Phi_2
$$

$$
E[X_2^Ry] = \Phi_1 + X_2^R\beta_2
$$

$$
E[Z'y] = Z^X_1\beta_1 + Z^X_2\beta_2,
$$

where $Z$ is any basis for the intersection of $C(X_1)$ and $C(X_2)$.

By letting $Z = M$, the row space of $(X_1, X_2)$ can be decomposed into the following subspaces:

$$
R(X_1^R, X_2^R) = R(X_1^R, X_2^R, \Phi)
$$

$$
R(X_2^R, X_1^R) = R(\Phi, X_1^R, X_2^R)
$$

$$
R(M, X_1, X_2) = R(MX_1, MX_2)
$$

where $\Phi$ is the null matrix of the appropriate sizes. The above row spaces characterize the vectors $\lambda$ such that $\lambda'\beta$ is estimable. From the dimension argument in the proof of Theorem 4.2, it follows that

$$
dim[R(X_1^R, X_2^R, \Phi)] = q_1 - q_{12},
$$

$$
dim[R(\Phi, X_1^R, X_2^R)] = q_2 - q_{12},
$$

$$
dim[R(MX_1, MX_2)] = q_{12}.
$$

Hence partitioning $C(X)$ induces a similar partition on $R(X)$. 
Vectors contained in $R(MX_1, MX_2)$ have a particularly unique interpretation in most experimental studies. They characterize the confounding in an experiment. This can be stated more concisely in the following definitions.

**Definition (Complete Confounding):** The functions $\lambda_1' \beta_1$ and $\lambda_2' \beta_2$ are said to be completely confounded if the BLUE, $a'y$, of their sum $\lambda_1' \beta_1 + \lambda_2' \beta_2$ exists such that $a \in C(x_1) \cap C(x_2)$.

**Definition (Separably Estimable):** The function $\lambda_1' \beta_1 + \lambda_2' \beta_2$ is said to be separably estimable if $\lambda_1' \beta_1 + \lambda_2' \beta_2$ is estimable and its BLUE, $a'y$, is such that $a$ can be written nontrivially as the sum of two vectors

$$a = a_1 + a_2$$

with $a_1 \in C(x_1) \cap C(x_2)$ and $a_2 \notin C(x_1) \cap C(x_2)$.

The term aliasing is derived from the fact that the vector $a$, when $a'y$ is BLUE for the completely confounded effect $\lambda_1' \beta_1 + \lambda_2' \beta_2$, can be represented as

$$a = x_1 \eta_1$$

or

$$a = x_2 \eta_2$$

for some vectors $\eta_1$ and $\eta_2$. That is, the vector $a$ has two distinct "names." However, caution should be exercised when using the term aliasing for the functions $\lambda_1' \beta_1$ and $\lambda_2' \beta_2$. The function $a'y$ is not BLUE for a function of $\beta_1$ or a function of $\beta_2$. The function $a'y$ is BLUE for the sum $\lambda_1' \beta_1 + \lambda_2' \beta_2$. 
C. Estimation In Submodels

Some relationships between estimation using the full model and estimation using submodels will now be explored. Consider the linear models:

\[(4.27)\]

Model (1) \[ y = X_1\beta_1 + X_2\beta_2 + \varepsilon_1 \]

Model (2) \[ y = X_1\beta_1 + \varepsilon_2 \]

Model (3) \[ y = X_2\beta_2 + \varepsilon_3 \]

where \(\varepsilon_1, \varepsilon_2, \varepsilon_3\) are vectors of independently distributed random variables with zero means and constant variances, \(\sigma_1^2, \sigma_2^2, \sigma_3^2\). Using this notation the following simple theorems can be established.

**Theorem 4.3** The set of estimable functions \(\lambda_1^\prime \beta_1\) that are estimable with Model (1) form a subset of those estimable with Model (2).

**Proof** If \(\lambda_1^\prime \beta_1\) is estimable with Model (1), there exists a vector \(a\) such that \(\lambda_1^\prime = a^\prime R_2 X_1\). This clearly implies that \(\lambda_1^\prime\) is in \(R(X_1)\), and therefore \(\lambda_1^\prime \beta_1\) is estimable with Model (2).

**Theorem 4.4** If the function \(a'y\) is BLUE for its expectation with Model (2), then the function \(a'R_2y\) is BLUE for its expectation with Model (1).

**Proof** The function \(a'y\) being BLUE for its expectation under Model (2) implies that there exists a vector \(\rho_1\) such that \(a = X_1 \rho_1\). Therefore, \(R_2a = R_2X_1 \rho_1\) which is sufficient for \(a'R_2y\) to be BLUE for its expectation under Model (1).

**Theorem 4.5** The BLUE's of estimable functions of \(\beta_1\) alone with Model (1) are also BLUE's of estimable functions of \(\beta_1\) with Model (2) if and only if \(C(X_1)\) forms an invariant subspace of the matrix \(R_2\).
Proof Assume $C(X_1)$ forms an invariant subspace of $R_2$, i.e. for any vector $a$ contained in $C(X_1)$, $R_2a$ is also contained in $C(X_1)$. Since any vector $a$ contained in $C(R_2X_1)$ can be written as $R_2b$ where $b$ is contained in $C(X_1)$, the vector $a$ must also be contained in $C(X_1)$. This implies that $a'y$ is BLUE for its expectation with Model (2).

Now assume that BLUE's of estimable functions of $\beta_1$ alone with Model (1) are also BLUE's of estimable functions with Model (2). Let $a'y$ be BLUE for its expectation with Model (2). Then by Theorem 4.4, $a'R_2y$ is BLUE for its expectation with Model (1). Since, by assumption, $a'R_2y$ is also BLUE of its expectation with Model (2), it follows that $R_2a \in C(X_1)$. Therefore, it has been demonstrated that for an arbitrary vector $a \in C(X_1)$, the vector $R_2a$ is also contained in $C(X_1)$, which is the definition of invariance.

Theorem 4.6 A subset of the BLUE's of estimable functions of $\beta_1$ alone with Model (1) is also a subset of the BLUE's of estimable functions with Model (2) if and only if a subspace of $C(X_1)$ forms an invariant subspace of the matrix $R_2$.

Proof The proof is identical to that of the previous theorem.

Theorem 4.7 If $a'y$ is BLUE for $\lambda_1\beta_1$ with Model (2) and $a'y$ is also BLUE for $\lambda_2\beta_2$ with Model (3), then $a'y$ is BLUE for $\lambda_1\beta_1 + \lambda_2\beta_2$ with Model (1).

Proof $a'y$ BLUE for $\lambda_1\beta_1$ with Model (2) implies $a = X_1\rho_1$ for some vector $\rho_1$. Similarly, there exists a vector $\rho_2$ such that $a = X_2\rho_2$. These two conditions imply that $a$ is a vector in the intersection of $C(X_1)$ and $C(X_2)$ and is therefore BLUE for its expectation.
Theorem 4.8 If the linear function $\lambda_1'\beta_1 + \lambda_2'\beta_2$ is estimable with Model (1), then $\lambda_1'\beta_1$ is estimable with Model (2) and $\lambda_2'\beta_2$ is estimable with Model (3).

Proof Let $a$ be such that $a'y$ is BLUE for $\lambda_1'\beta_1 + \lambda_2'\beta_2$. Then

$$E_1[a'y] = a'X_1\beta_1 + a'X_2\beta_2,$$

$$E_2[a'y] = a'X_1\beta_1,$$

and

$$E_3[a'y] = a'X_2\beta_2,$$

where the subscript on the expectation indicates the model used. From the above expectations it follows that $\lambda_1 \in R(X_1)$ and $\lambda_2 \in R(X_2)$.

The results of Theorem 4.5 suggest a correspondence between that theorem and the results of Zyskind (1967) concerning the equality of best and simple least squares estimators. If $C(X_1)$ is contained in $C(R_2)$, then the reduced normal equations for eliminating $\beta_2$,

$$(4.28) \quad X_1'R_2X_1\beta_1 = X_1'R_2y,$$

can be thought of as general normal equations for fitting the model

$$(4.29) \quad R_2y = z = X_1\beta_1 + \varepsilon,$$

where $\varepsilon$ has variance covariance matrix $\sigma^2R_2$. Since the generalized inverse of $R_2$ is equal to $R_2$, Corollary 1.1 of Zyskind and Martin (1969) implies that with the conditions above, possible normal equations are
equations 4.28. Therefore, the following conditions from Zyskind (1967) establish the equivalence between BLUE's of estimable functions of $\beta_1$ alone with Model (1) and BLUE's of estimable functions of $\beta_1$ with Model (2).

i) A subset of $q_1$ eigenvectors of $R_1$ exists forming a basis for $C(X_1)$, where $q_1 = \text{rank}(X_1)$.

ii) A full rank reparametrization exists so that $E(y) = X_1\beta_1 = W\Theta$, where the columns of the $n \times q_1$ matrix are mutually orthogonal eigenvectors of $R_2$.

iii) The matrix $R_2$ is expressible in the form

$$R_2 = \begin{pmatrix} 0_1 \end{pmatrix} \begin{pmatrix} 0_1' & 0_0' \end{pmatrix} D \begin{pmatrix} 0_1 \\ 0_0 \end{pmatrix},$$

where the matrix $0' = (0_1, 0_0)$ is orthogonal and $0_1'$ is any orthonormal basis of $X_1$, $0'_0$ is any orthonormal basis of the orthogonal complement of the column space of $X_1$, and $D$ is any diagonal matrix with nonnegative elements.

iv) The matrix $R_2$ can be diagonalized by an orthogonal matrix specified in iii).

v) If $M_1$ denotes the orthogonal matrix projection operator on the column space of $X_1$, then $R_2M_1 = M_1R_2$ is a relation which holds if and only if $R_2M_1$ is symmetric.

vi) A matrix $Q$ exists which satisfies the relation $R_2X_1 = X_1Q$.

Relationships similar to those above exist establishing the same correspondences between Model (1) and Model (3).
D. Decomposition of Sums of Squares

Thus far we have developed a decomposition of estimable functions and their corresponding BLUE's. Since we can only test hypotheses concerning functions that are estimable, it seems reasonable now to pursue possible partitions of the regression sum of squares from fitting the partitioned linear model.

The regression sum of squares can be written as

\[(4.30)\quad \text{SS(Regression)} = \hat{\beta}'X'X\hat{\beta},\]

or according to our partition,

\[(4.31)\quad \text{SS(Regression)} = (\hat{\beta}'_1,\hat{\beta}'_2) \begin{pmatrix} X'_1X'_1 & X'_1X'_2 \\ X'_2X'_1 & X'_2X'_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}'_1 \\ \hat{\beta}'_2 \end{pmatrix} = \hat{\beta}'_1X'_1X'_1\hat{\beta}'_1 + 2\hat{\beta}'_1X'_1X'_2\hat{\beta}'_2 + \hat{\beta}'_2X'_2X'_2\hat{\beta}'_2.\]

Also, since \(X'X\hat{\beta} = X'y\),

\[(4.32)\quad \text{SS(Regression)} = \hat{\beta}'_1X'_1y + \hat{\beta}'_2X'_2y.\]

Since the projection operators \(M, M_1,\) and \(M_2\) are used considerably in the next few sections, we will prove the following two theorems.

**Theorem 4.9** Let \(M\) denote the orthogonal projection operator on \(C(X_1) \cap C(X_2)\). Then the matrix expression

\[M_1M_2 = M_2M_1 = M\]
holds if and only if

\[ M_{1}X_{2} = MX_{2} \]

or

\[ M_{2}X_{1} = MX_{1} . \]

**Proof**  Assume \( M_{1}M_{2} = M_{2}M_{1} = M \). Since \( X_{1} = M_{1}X_{1} \) and \( X_{2} = M_{2}X_{2} \),

\[ M_{2}X_{1} = M_{2}M_{1}X_{1} = MX_{1} \]

and

\[ M_{1}X_{2} = M_{1}M_{2}X_{2} = MX_{2} . \]

Assume \( M_{2}X_{1} = MX_{1} \) or \( M_{1}X_{2} = MX_{2} \). Postmultiplying the first expression by \((X_{1}X_{1})^{-1}X_{1}'\) gives

\[ M_{2}M_{1} = M . \]

Postmultiplying the second expression by \((X_{2}X_{2})^{-1}X_{2}'\) gives

\[ M_{1}M_{2} = M . \]

**Theorem 4.10**  The matrix \( X_{1}'(I - M)X_{2} \) is the zero matrix if and only if \( M_{1}M_{2} = M_{2}M_{1} = M \).

**Proof**  Assume \( M_{1}M_{2} = M_{2}M_{1} = M \). Since \( X_{1} = M_{1}X_{1} \) and \( X_{2} = M_{2}X_{2} \),
\[ X_1'(I - M)X_2 = X_1'M_1(I - M)M_2X_2 \]
\[ = X_1'(M_1M_2 - M)X_2 \]
\[ = X_1'(M - M)X_2 \]
\[ = \phi . \]

Assume \( X_1'(I - M)X_2 = \phi \), then premultiplying this expression by \( X_1(X_1'X_1)^{-1} \) and postmultiplying by \((X_2'X_2)^{-1}X_2^t \) gives
\[ M_1(I - M)M_2 = M_1M_2 - M = \phi . \]

Therefore, since \( M \) is symmetric,
\[ M_1M_2 = M = M_2M_1 . \]

Now consider the following decomposition,

\[
\begin{align*}
\text{SS(Regression)} &= (\hat{\beta}_1, \hat{\beta}_2) \begin{pmatrix} X_1'M_1X_1 & X_1'M_2X_2 \\ X_2'M_1X_1 & X_2'M_2X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \\
&\quad + (\hat{\beta}_1, \hat{\beta}_2) \begin{pmatrix} X_1'(I - M)X_1 & X_1'(I - M)X_2 \\ X_2'(I - M)X_1 & X_2'(I - M)X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} .
\end{align*}
\]

From 4.32, it is clear that the two terms of the above decomposition can also be written as

\[
\begin{align*}
\text{SS(Regression)} &= y'M(\hat{\beta}_1 + X_2\hat{\beta}_2) + y'(I - M)(\hat{\beta}_1 + X_2\hat{\beta}_2) .
\end{align*}
\]
The first term gives

\[(4.35) \quad y'M(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) = y'M M_1(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) = y'M X_1(X_1'X_1)^{-1}X_1'(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) .\]

Since \(X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 = X_1'y,\)

\[y'M(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) = y'M M_1 y = y'M y .\]

This can be thought of as the sum of squares removed by the completely confounded effects. Since the spaces \(C(R_2X_1)\) and \(C(R_1X_2)\) are orthogonal to \(C(M)\), the second term, \(y'(I - M)(X_1\hat{\beta}_1 + X_2\hat{\beta}_2)\), should contain the information concerning estimable functions of \(\beta_1\) alone and estimable functions of \(\beta_2\) alone. First, the information concerning \(\beta_2\) alone will be removed by considering the reduced normal equations for \(\beta_2\) eliminating \(\beta_1\),

\[(4.36) \quad X_1'R_2X_2\hat{\beta}_2 = X_1'R_1y .\]

This yields a corresponding \(\hat{\beta}_1\) as any solution to

\[(4.37) \quad X_1'X_1\hat{\beta}_1 = X_1'(y - X_2\hat{\beta}_2) ,\]

and the sum of squares for \(\beta_2\) eliminating \(\beta_1\) as
Subtracting this sum of squares from the second term in 4.34, we obtain

\[ y'(I - M)(X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2) - \hat{\beta}_2 \hat{X}_1 \hat{X}_2 \hat{\beta}_2 \]

\[ = y'(I - M)(X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2) - y'R \hat{X}_2 \hat{\beta}_2 \]

\[ = y'(I - M)X_1 \hat{\beta}_1 + y'(I - M - R_1)X_2 \hat{\beta}_2 \]

\[ = y'(I - M)X_1 \hat{\beta}_1 + y'(M_1 - M)X_2 \hat{\beta}_2 \]

\[ = y'(M_1 - M)(X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2) . \]

Since \( M_1 M = M M_1 \) and \( M_1 (X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2) = M_1 y \), the expression above reduces to

\[ y'(M_1 - M)y . \]

We will now show that \( \hat{\beta}_1 \), defined by 4.37, satisfies the reduced normal equations for \( \beta_1 \) eliminating \( \beta_2 \).

\[ X_1' R_2 X_1 \hat{\beta}_1 = X_1' X_1 \hat{\beta}_1 - X_1' M_2 X_1 \hat{\beta}_1 \]

\[ = X_1' y - X_1' X_2 \hat{\beta}_2 - X_1' M_2 X_1 \hat{\beta}_1 \]

\[ = X_1' y - X_1' M_2 (X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2) \]

\[ = X_1' y - X_1' M_2 y \]

\[ = X_1' R_2 y . \]
Therefore, the sum of squares for $\beta_1$ eliminating $\beta_2$ can be written as

\[(4.42) \quad SS(\beta_1 \text{ eliminating } \beta_2) = \hat{\beta}_1' X_1' R_{12} X_1 \hat{\beta}_1.\]

Subtracting this expression from the second term in 4.34 we obtain a difference similar to 4.40,

\[(4.43) \quad y'(M_2 - M)y.\]

Table 4.1 contains this information displayed in an analysis of variance format where $q_1 = \text{rank}(X_1)$, $q_2 = \text{rank}(X_2)$, and $q_{12} = \text{dimension of the intersection of } C(X_1) \text{ and } C(X_2)$.

There are several interesting properties of this decomposition that can be noted from the Analysis of Variance table.

i) The sum of lines (1) and (3) give $y'M_1y$ or $y'M_2y$. These are $SS(\beta_1 \text{ ignoring } \beta_2)$ and $SS(\beta_2 \text{ ignoring } \beta_1)$, respectively.

ii) Since the matrix of the quadratic form in line (3), $M_1 - M$ or $M_2 - M$, is symmetric and idempotent, it follows that $y'(M_1 - M)y$ and $y'(M_2 - M)y$ are always nonnegative. This implies that the sum of squares for $\beta_i$ eliminating $\beta_j$ is always less than the sum of squares for regression minus the sum of squares for confounding.

iii) The sums of squares in lines (1), (2) and (3) can be written as $y'My$, $y'R_{i_j} (X_1' R_{i_j})^{-1} X_1' R_{i_j} y$, and $y'(M_j - M)y$ for $i, j = 1, 2$. Letting $R_{i_j} (X_1' R_{i_j})^{-1} X_1' R_{i_j} = M_{ij}$, and $P$ be the orthogonal projection operator on $C(X_1, X_2)$, we have,
Table 4.1 The analysis of variance for $\beta_j$ eliminating $\beta_j$.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Source</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) SS(completely confounded effects)</td>
<td>$q_{12}$</td>
<td>$y'M'y$</td>
<td>SS(completely confounded effects)</td>
<td>$q_{12}$</td>
<td>$y'M'y$</td>
</tr>
<tr>
<td>(2) SS($\beta_2$ eliminating $\beta_1$)</td>
<td>$q_2-q_{12}$</td>
<td>$\hat{\beta}'X'R_1X_2\hat{\beta}$</td>
<td>SS($\beta_1$ eliminating $\beta_2$)</td>
<td>$q_1-q_{12}$</td>
<td>$\hat{\beta}'X'R_2X_1\hat{\beta}$</td>
</tr>
<tr>
<td>(3) Difference (4)-(1)-(2)</td>
<td>$q_1-q_{12}$</td>
<td>$y'(M_1-M)y$</td>
<td>Difference (4)-(1)-(2)</td>
<td>$q_2-q_{12}$</td>
<td>$y'(M_2-M)y$</td>
</tr>
<tr>
<td>(4) Total = SS(Regression)</td>
<td>$q_1+q_2-q_{12}$</td>
<td>$\hat{\beta}'X'\hat{\beta}$</td>
<td>Total = SS(Regression)</td>
<td>$q_1+q_2-q_{12}$</td>
<td>$\hat{\beta}'X'\hat{\beta}$</td>
</tr>
</tbody>
</table>

$\hat{\beta}$ satisfies $X'\hat{\beta} = X'y$
\[ I_n = (I_n - P) + M + M_{1j} + (M_j - M) = A_1 + A_2 + A_3 + A_4, \]

where

1) \( A_i A_j = 0 \) for \( i \neq j \).

2) Each \( A_i \) is idempotent

3) \( \sum_i \text{rank}(A_i) = n \).

The above properties, with the additional property,

4) There exists an orthogonal matrix \( Q \) which diagonalizes simultaneously the matrices \( A_i \) in such a manner that each diagonal matrix \( QA_i Q' \) has \( r_i = \text{rank}(A_i) \) consecutive diagonal elements equal to unity and the remaining \( n - r_i \) diagonal elements equal to zero.

constitute the familiar, more general form of Cochran's theorem.

Expression 4.44 suggests the following decompositions of \( P \), the orthogonal projection operator on \( C(X_1, X_2) \),

(4.45) (1) \( P = M + (M_1 - M) + M_{21} \),

(2) \( P = M + (M_2 - M) + M_{12} \).

Since \( (X_1, X_2) = P(X_1, X_2) \), expression 4.45 leads to the following decompositions of \( (X_1, X_2) \),
(1) \[(X_1, X_2) = (MX_1, MX_2) + ((I-M)X_1, (M_1-I)X_2) + (M_{21}X_1, M_{21}X_2)\]

(4.46)

(2) \[(X_1, X_2) = (MX_1, MX_2) + ((M_2-I)X_1, (I-M)X_2) + (M_{12}X_1, M_{12}X_2)\]

Since the orthogonal projection operator $M_{21}$, for example, can be written as

\[
M_{21} = R_2'(X_2'R_2'X_2'R_2')^{-1}X_2'R_2',
\]

we have $M_{21}X_1 = \phi$ and $M_{21}X_2 = R_1X_2$. This gives

(1) \[(X_1, X_2) = (MX_1, MX_2) + ((I-M)X_1, (I-M)X_2) + (\phi, R_1X_2)\]

(4.47)

(2) \[(X_1, X_2) = (MX_1, MX_2) + ((I-M)M_2X_1, (I-M)X_2) + (R_2X_1, \phi)\].

In order to examine some joint properties of the two representations of $(X_1, X_2)$, consider the following composite decomposition,

(4.48) \[(X_1, X_2) = (MX_1, MX_2) + ((M_2-I)X_1, (M_1-I)X_2) + (R_2X_1, \phi) + (\phi, R_1X_2)\].

This gives the corresponding decomposition of $X'X$. 
Therefore, the regression sum of squares can be written as

\[
(4.50) \quad \hat{\beta}'X'X\hat{\beta} = (\hat{\beta}_1'X_1' + \hat{\beta}_2'X_2')M(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) \\
+ (\hat{\beta}_1'X_1' + \hat{\beta}_2'X_2') R_2(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) \\
+ (\hat{\beta}_1'X_1' + \hat{\beta}_2'X_2') R_1(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) \\
+ (\hat{\beta}_1'X_1' + \hat{\beta}_2'X_2')(I-M-R_1-R_2)(X_1\hat{\beta}_1 + X_2\hat{\beta}_2) .
\]

Since \(X\hat{\beta}\) is invariant for all solutions \(\hat{\beta}\) to the normal equations, each of the terms in the above expression is also invariant for all solutions \(\hat{\beta}\).

The analysis of variance for this decomposition is presented in Table 4.2. An extension of this decomposition will be introduced in subsequent chapters and used to construct the more standard analyses.
To discuss tests of hypotheses, one must examine the expectation of the above sums of squares and their corresponding mean squares, assuming the parameters satisfy certain properties. We will briefly examine expressions for expected mean squares under the random effects model, fixed effects model, and mixed effects model hypotheses. In the development that follows, the random variables introduced are assumed to come from an infinite population. In some cases that may not be all together appropriate and some form of randomization technique such as that discussed by Kempthorne (1955) and Zyskind (1962) should be used. However, an ultimate goal is to apply the theory developed here to a computational procedure. The "infinite" assumption above provides expressions for expected mean squares that involve quantities readily available from other calculations.
In a random effects model, the null hypothesis states that $\beta_1$ and $\beta_2$ are independent vectors of random variables having means zero and variances $\sigma_1^2 I$ and $\sigma_2^2 I$, respectively. Under these conditions the vector $y$ has expectation zero and variance given by

$$\text{(4.51)} \quad \text{Var}(y) = \sigma_1^2 I + \sigma_2^2 I + \sigma_1^2 X_1' X_1 + \sigma_2^2 X_2' X_2.$$ 

It is well known that if the vector random variable $y$ has mean $\mu$ and variance-covariance matrix $\Sigma$, then

$$\text{(4.52)} \quad \text{E}[y'y] = \mu' \Sigma \mu + \text{trace}(A \Sigma).$$

In subsequent expressions, trace($A$) will be abbreviated $\text{tr}(A)$.

Using the variance-covariance matrix defined in 4.51, the expectation of $y'y$ can be written as

$$\text{(4.53)} \quad \text{E}[y'y] = \sigma_1^2 \text{tr}(A) + \sigma_1^2 \text{tr}(X_1' A X_1) + \sigma_2^2 \text{tr}(X_2' A X_2).$$

This expression will be used to derive the expectations of the sums of squares in Table 4.1. The expectation of the sum of squares for completely confounded effects can be written as

$$\text{(4.54)} \quad \text{E}[y'y] = q_{12} \sigma_1^2 + \sigma_1^2 \text{tr}(X_1' A X_1) + \sigma_2^2 \text{tr}(X_2' A X_2).$$

This gives the expected mean square for completely confounded effects [abbreviated EMS(confounding)] as

$$\text{(4.55)} \quad \text{EMS(confounding)} = \sigma_1^2 + \frac{\sigma_1^2 \text{tr}(X_1' A X_1)}{q_{12}} + \frac{\sigma_2^2 \text{tr}(X_2' A X_2)}{q_{12}}.$$
Because of the nature of the solutions $\tilde{\beta}_2$ to the reduced normal equations for $\beta_2$ eliminating $\beta_1$, the $\text{SS}(\beta_2$ eliminating $\beta_1)$ can be written as

$$
(4.56) \quad \tilde{\beta}_2'X'X\tilde{\beta}_2 = y'R_1X_1(\tilde{X}_2'X_2)^{-1}\tilde{X}_2'R_2y.
$$

The expected value of this sum of squares and corresponding mean square are

$$
(4.57) \quad E[\text{SS}(\beta_2\text{ eliminating } \beta_1)] = \sigma^2(q_2-q_{12}) + \sigma_2^2\text{tr}(X_2'X_2),
$$

and

$$
(4.58) \quad \text{EMS}(\beta_2\text{ eliminating } \beta_1) = \sigma^2 + \sigma_2^2\text{tr}(X_2'X_2)/(q_2-q_{12}).
$$

Similarly, for $\beta_1$ eliminating $\beta_2$,

$$
(4.59) \quad E[\text{SS}(\beta_1\text{ eliminating } \beta_2)] = \sigma^2(q_1-q_{12}) + \sigma_1^2\text{tr}(X_1'X_1),
$$

and

$$
(4.60) \quad \text{EMS}(\beta_1\text{ eliminating } \beta_2) = \sigma^2 + \sigma_1^2\text{tr}(X_1'X_1)/(q_1-q_{12}).
$$

The differences in line (3) of Table 4.1 have expectations and expected mean squares given by:
(4.61) \[ E[y'(M_1-M)y] = \]
\[ \sigma^2(q_1-q_{12}) + \sigma_1^2 \text{tr}(X_1'(M_1-M)X_1) + \sigma_2^2 \text{tr}(X_2'(M_1-M)X_2), \]

(4.62) \[ \text{EMS}[y'(M_1-M)y] = \]
\[ \sigma^2 + \sigma_1^2 \text{tr}(X_1'(M_1-M)X_1)/(q_1-q_{12}) + \sigma_2^2 \text{tr}(X_2'(M_1-M)X_2)/(q_1-q_{12}), \]

(4.63) \[ E[y'(M_2-M)y] = \]
\[ \sigma^2(q_2-q_{12}) + \sigma_1^2 \text{tr}(X_1'(M_2-M)X_1) + \sigma_2^2 \text{tr}(X_2'(M_2-M)X_2)/(q_2-q_{12}), \]

and

(4.64) \[ \text{EMS}[y'(M_2-M)y] = \]
\[ \sigma^2 + \sigma_1^2 \text{tr}(X_1'(M_2-M)X_1)/(q_1-q_{12}) + \sigma_2^2 \text{tr}(X_2'(M_2-M)X_2)/(q_1-q_{12}). \]

These expressions are displayed in Table 4.3.

For the fixed effects model, the null hypothesis states that \( \beta_1 \) and \( \beta_2 \) are fixed vectors of constants, say \( b_1 \) and \( b_2 \). Then

(4.65) \[ E[y] = X_1b_1 + X_2b_2, \]

and

(4.66) \[ E[yy'] = (X_1b_1 + X_2b_2)(b_1'X_1 + b_2'X_2) + \sigma^2 I. \]
Table 4.3. Expected mean squares for $\beta_i$ eliminating $\beta_j$.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding</td>
<td>$q_{12}$</td>
<td>$\sigma^2 + \frac{\sigma^2 \text{tr}(X_1'MX_1)}{q_{12}} + \frac{\sigma^2 \text{tr}(X_2'MX_2)}{q_{12}}$</td>
</tr>
<tr>
<td>$\beta_1$ eliminating $\beta_2$</td>
<td>$q_1-q_{12}$</td>
<td>$\sigma^2 + \frac{\sigma^2 \text{tr}(X_1'R_2X_1)}{(q_1-q_{12})}$</td>
</tr>
<tr>
<td>Remainder</td>
<td>$q_2-q_{12}$</td>
<td>$\sigma^2 + \frac{\sigma^2 \text{tr}(X_1(M_2-M)X_1)}{(q_1-q_{12})}$ + $\frac{\sigma^2 \text{tr}(X_2(M_2-M)X_2)}{(q_2-q_{12})}$</td>
</tr>
</tbody>
</table>

| Confounding                 | $q_{12}$           | $\sigma^2 + \frac{\sigma^2 \text{tr}(X_1'MX_1)}{q_{12}} + \frac{\sigma^2 \text{tr}(X_2'MX_2)}{q_{12}}$ |
| $\beta_2$ eliminating $\beta_1$ | $q_2-q_{12}$      | $\sigma^2 + \frac{\sigma^2 \text{tr}(X_2'R_1X_2)}{(q_2-q_{12})}$                     |
| Remainder                   | $q_1-q_{12}$       | $\sigma^2 + \frac{\sigma^2 \text{tr}(X_1(M_1-M)X_1)}{(q_1-q_{12})}$ + $\frac{\sigma^2 \text{tr}(X_2(M_1-M)X_2)}{(q_2-q_{12})}$ |
The expectation of the quadratic form $y'Ay$ is then

\[(4.67) \quad E[y'Ay] = \sigma^2 \text{tr}(A) + (b'_1X'_1 + b'_2X'_2) A (X_{1b1} + X_{2b2}).\]

Thus, the expectation of the sum of squares and expected mean square for completely confounded effects can be written as

\[(4.68) \quad E[SS(\text{confounding})] = q_{12} \sigma^2 + b'_1X'_1X_{1b1} + b'_2X'_2X_{2b2} + 2b'_1X'_1X_{1b2},\]

and

\[(4.69) \quad \text{EMS}(\text{confounding}) = \sigma^2 + (b'_1X'_1 + b'_2X'_2) M (X_{1b1} + X_{2b2})/q_{12}.\]

The expected mean squares for the quantities in Table 4.1 are found in Table 4.4.
Table 4.4. Expected mean squares for a fixed effects model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding</td>
<td>$q_{12}$</td>
<td>$\sigma^2 + (X_1b_1 + X_2b_2)'M(X_1b_1 + X_2b_2)/q_{12}$</td>
</tr>
<tr>
<td>$\beta_1$ eliminating $\beta_2$</td>
<td>$q_{1}-q_{12}$</td>
<td>$\sigma^2 + b_1'X_1R_1X_1b_1/(q_{1}-q_{12})$</td>
</tr>
<tr>
<td>Remainder</td>
<td>$q_{2}-q_{12}$</td>
<td>$\sigma^2 + (X_1b_1 + X_2b_2)'(M_2-M)(X_1b_1 + X_2b_2)/(q_{2}-q_{12})$</td>
</tr>
<tr>
<td>Confounding</td>
<td>$q_{12}$</td>
<td>$\sigma^2 + (X_1b_1 + X_2b_2)'M(X_1b_1 + X_2b_2)/q_{12}$</td>
</tr>
<tr>
<td>$\beta_2$ eliminating $\beta_1$</td>
<td>$q_{2}-q_{12}$</td>
<td>$\sigma^2 + b_2'X_2R_2X_2b_2/(q_{2}-q_{12})$</td>
</tr>
<tr>
<td>Remainder</td>
<td>$q_{1}-q_{12}$</td>
<td>$\sigma^2 + (X_1b_1 + X_2b_2)'(M_1-M)(X_1b_1 + X_2b_2)/(q_{1}-q_{12})$</td>
</tr>
</tbody>
</table>

To complete the development of expressions for expected mean squares, assume one of the sets of parameters is random and the other is fixed. This situation corresponds to the mixed effects model. Since the problem is symmetric in $\beta_1$ and $\beta_2$, assume $\beta_1$ is random with mean zero and variance $\sigma^2_1I$ and $\beta_2$ is a vector of fixed constants $b_2$. The expected value of $yy'$ is
This gives the following expectation for the quadratic form $y' Ay$,

\[(4.71) \quad E[y'y] = \sigma^2 \text{tr}(A) + \sigma_1^2 \text{tr}(X_1'MX_1) + b_2'X_2b_2 \cdot \]

Applying this result to the expressions in Table 4.1 gives the quantities in Table 4.5.

**Table 4.5. Expected mean squares for the mixed model.**

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounding</td>
<td>$q_{12}$</td>
<td>$\sigma^2 + \sigma_1^2 \text{tr}(X_1'MX_1)/q_{12} + b_2'X_2b_2/q_2$</td>
</tr>
<tr>
<td>$\beta_1$ eliminating $\beta_2$</td>
<td>$q_1 - q_{12}$</td>
<td>$\sigma^2 + \sigma_1^2 \text{tr}(X_1'MX_1)/(q_1 - q_{12})$</td>
</tr>
</tbody>
</table>
| Remainder                   | $q_2 - q_{12}$     | $\sigma^2 + \sigma_1^2 \text{tr}(X_1'(M_2 - M)X_1)/(q_2 - q_{12})$  \
|                            |                    | + $b_2'X_2'(M_2 - M)X_2b_2/(q_2 - q_{12})$                                          |
| Confounding                 | $q_{12}$           | $\sigma^2 + \sigma_1^2 \text{tr}(X_1'MX_1)/q_{12} + b_2'X_2b_2/q_{12}$             |
| $\beta_2$ eliminating $\beta_1$ | $q_2 - q_{12}$    | $\sigma^2 + b_2'X_2'R_2X_2b_2/(q_2 - q_{12})$                                      |
| Remainder                   | $q_1 - q_{12}$     | $\sigma^2 + \sigma_1^2 \text{tr}(X_1'(M_2 - M)X_1)/(q_1 - q_{12})$  \
|                            |                    | + $b_2'X_2'(M_2 - M)X_2b_2/(q_1 - q_{12})$                                          |
It is worth noting that several but not all of the quadratic forms presented in these tables are independent. This is due to the fact that the product \( R_1R_2 \) is not in general the null matrix. The mean squares calculated in the analysis of \( \beta_i \) eliminating \( \beta_j \) are independently distributed and therefore the \( F \) statistic can be used for testing hypotheses. However, the quadratic forms calculated for the analysis of \( \beta_i \) eliminating \( \beta_j \) are not independent of those calculated for \( \beta_j \) eliminating \( \beta_i \).

E. The Reduced Normal Equations

One of the most difficult problems encountered when working with a general partitioned linear model is the complexity of the reduced normal equations for \( \beta_i \) eliminating \( \beta_j \),

\[
(4.72) \quad X'_{iJ}X_{ii} = X'_{iJ}y .
\]

In this section some methods for obtaining solutions to equations 4.75 and their properties are discussed.

Recall from Chapter II that Wilkinson (1970) introduces a shrinkage operator \( Q \) defined by \( Q = R_1M_2R_1 \). Using the reduced minimal polynomial \( p(x) \) of \( Q \), he calculates a conditional inverse \( C \) of \( X'_{21}X_2 \) by

\[
(4.73) \quad C = (X'_{21}X_2)^{-q[x'_{21}X_2(x'_{21}X_2)^{-}]},
\]

where \( (X'_{21}X_2)^{-} \) is any conditional inverse of \( X'_{21}X_2 \), and \( q(x) \) is a polynomial defined by

\[
(4.74) \quad q(x) = [1 - p(x)]/x .
\]
This gives the corresponding solutions $\tilde{\beta}_2$ to the equations

\begin{equation}
(4.75) \quad X_2'X_2\tilde{\beta}_2 = X_2'R_1y,
\end{equation}

as

\begin{equation}
(4.76) \quad \tilde{\beta}_2 = (X_2'X_2)^{-q}[X_2'R_1X_2(X_2'X_2)^{-}]X_2'R_1y.
\end{equation}

Equation 4.76 suggests that any solution to the equations

\begin{equation}
(4.77) \quad X_2'X_2\beta_2 = q[X_2'R_1X_2(X_2'X_2)^{-}]X_2'R_1y
\end{equation}

is also a solution to equations 4.75. This can easily be shown using the following results.

**Theorem 4.11** The matrices

i) $R_1M_2R_1$,

ii) $M_2R_1$,

iii) $R_1M_2$,

iv) $(X_2'X_2)^{-}X_2'R_1X_2$, and

v) $X_2'R_1X_2(X_2'X_2)^{-}$

all have the same reduced minimal polynomial. James and Wilkinson (1971) proved results iv) and v).

**Proof** Let $p(x)$ be the reduced minimal polynomial of $R_1M_2R_1$. Then $p(x)$ is the polynomial of minimal degree such that

\begin{equation}
(4.78) \quad R_1M_2R_1 p(R_1M_2R_1) = \phi.
\end{equation}
Since \( p(x) \) is a polynomial and the matrices \( R_1 \) and \( M_2 \) are both symmetric and idempotent, it follows that

\[
(4.79) \quad R_1M_2R_1 \quad p(R_1M_2R_1) = R_1M_2R_1 \quad p(M_2R_1) = R_1M_2R_1 \quad p(M_2R_1).
\]

Multiplying both sides of 4.79 by \( p(R_1M_2) \) gives

\[
(4.80) \quad p(R_1M_2)R_1M_2R_1 \quad p(M_2R_1) = [M_2R_1 \quad p(M_2R_1)]'[M_2R_1 \quad p(M_2R_1)]
= \phi.
\]

Therefore, \( M_2R_1 \quad p(M_2R_1) = \phi \) which implies that the degree of the reduced minimal polynomial of \( M_2R_1 \) is not greater than the degree of \( p(x) \).

Let \( p'(x) \) denote the reduced minimal polynomial of \( M_2R_1 \). Then

\[
(4.81) \quad M_2R_1 \quad p'(M_2R_1) = \phi.
\]

Multiplying both sides of 4.81 by \( R_1 \) gives

\[
(4.82) \quad R_1M_2R_1 \quad p'(M_2R_1) = \phi.
\]

Since \( R_1 \) is idempotent and \( p'(x) \) is a polynomial,
Therefore, the degree of the reduced minimal polynomial of $M_2R_1$ is not less than the degree of $p(x)$. Since the reduced minimal polynomial of a matrix is unique except for a scalar multiplier, the two polynomials $p'(x)$ and $p(x)$ are identical. From the fact that the reduced minimal polynomial of a matrix and the reduced minimal polynomial of its transpose are the same, it follows that the matrices $M_2R_1$ and $R_1M_2$ have the same reduced minimal polynomial.

Let $p'(x)$ denote the reduced minimal polynomial of $X_2'X_2(X_2'X_2)^-$. This implies

\begin{equation}
X_2'X_2(X_2'X_2)^-p'[X_2'X_2(X_2'X_2)^-] = \phi.
\end{equation}

Multiplying the above expression on the left by $X_2(X_2'X_2)^-$ and on the right by $X_2'$ gives

\begin{equation}
M_2R_1M_2 p'[R_1M_2] = \phi.
\end{equation}

Premultiplying by $p'(M_2R_1)$ gives

\begin{equation}
[p'(M_2R_1)M_2R_1][R_1M_2 p'(R_1M_2)] = \phi.
\end{equation}

Therefore the degree of the reduced minimal polynomial of $R_1M_2$ is not less than the degree of the reduced minimal polynomial of $X_2'X_2(X_2'X_2)^-$. 

\begin{equation}
R_1M_2 p'(M_2R_1) = R_1M_2 p'(R_1M_2)
= \phi.
\end{equation}
Let \( p(x) \) denote the reduced minimal polynomial of \( M_2R_1 \). Then,

\[
(4.87) \quad M_2R_1 p(M_2R_1) = \phi .
\]

Premultiplying by \( X_2^I \) and postmultiplying by \( X_2(X_2^I X_2)^- \) gives

\[
(4.88) \quad X_2^I R_1 p(M_2R_1) X_2(X_2^I X_2)^- = \phi .
\]

Since \( p(x) \) is a polynomial, expression 4.88 can be written

\[
(4.89) \quad X_2^I R_1 X_2(X_2^I X_2)^- p[X_2^I R_1 X_2(X_2^I X_2)^-] = \phi .
\]

Therefore, the degree of the reduced minimal polynomial of \( R_1 M_2 \) is not greater than the degree of the reduced minimal polynomial of \( X_2^I R_1 X_2(X_2^I X_2)^- \) and hence the two reduced minimal polynomials are the same.

The following theorem concerning the polynomial \( q(x) = \frac{1 - p(x)}{x} \) introduces a useful result. James and Wilkinson (1971) establish Theorem 4.12 for a specific matrix.

**Theorem 4.12** Let \( p(x) \) denote the reduced minimal polynomial of the square matrix \( A \). If \( p(x) \) has the form

\[
(4.90) \quad p(x) = 1 + a_1 x + \ldots + a_k x^k ,
\]

then there exists a polynomial \( q(x) \) such that \( q(A) \) is a conditional inverse of \( A \).

**Proof** Define

\[
(4.91) \quad q(x) = \frac{1 - p(x)}{x} .
\]
Then \( Aq(A)A = A(I - p(A)) = (I - p(A))A = A \). It follows that \( q(x) \) is the polynomial of minimal degree such that Theorem 4.12 holds. If \( A \) is nonsingular then \( p(A) = \phi \), and it follows that \( q(A) = A^{-1} \).

From Perlis (1952), page 186, it follows that for any symmetric matrix \( A \), the reduced minimal polynomial of \( A \) can be written as in 4.90. Therefore, the reduced minimal polynomial of the matrices discussed in Theorem 4.11 can be written as in 4.90.

Expression 4.77 can be written as

\[(4.92) \quad x_2^2 x_2 = x_2 q(R_1 M_2) R_1 y .\]

Letting \( A^+ \) be the conditional inverse of \( A \) defined by Theorem 4.12, the above expression becomes

\[(4.93) \quad x_2^2 x_2 = x_2 (R_1 M_2)^+ R_1 y .\]

Since \( x_2^2 = x_2 M_2 \) and because \( q(x) \) is a polynomial, these equations can also be written as

\[(4.94) \quad x_2^2 x_2 = x_2 (R_1 M_2)^+ R_1 y = x_2 x_2 M_2 R_1 (M_2 R_1)^+ y = x_2 (M_2 R_1)^+ M_2 R_1 y = x_2 x_2 (R_1 M_2 R_1)^+ y .\]
The predicted values $X_2\hat{\beta}_2$ from equations 4.94 can be written as

(4.95) \[ X_2\hat{\beta}_2 = (M_2 R_1)^+ M_2 R_1 y. \]

Therefore $X_2^1 R_1 X_2\hat{\beta}_2$, where $\hat{\beta}_2$ is any solution to 4.94, can be written as

(4.96) \[ X_2^1 R_1 X_2\hat{\beta}_2 = X_2^1 R_1 (M_2 R_1)^+ M_2 R_1 y = X_2^1 M_2 R_1 (M_2 R_1)^+ M_2 R_1 y = X_2^1 M_2 R_1 y = X_2^1 R_1 y. \]

Since the matrix $M_2 R_1 (M_2 R_1)^+$ is idempotent, it is a projection operator. The image of the projection is clearly $C(M_2 R_1)$.

It was demonstrated earlier that any solution $\hat{\beta}_2$ to the equations

(4.97) \[ X_2^1 X_2\hat{\beta}_2 = X_2^1 [I - X_1 (X_1^1 R_2 X_1)^- X_1^1 R_2] y \]

also satisfies equations 4.75.

It was also shown earlier (Theorem 3.1), when $M_1 M_2 = M$, that solutions to equations

(4.98) \[ X_2^1 X_2 \beta_2 = X_2^1 (I - M) y \]

also satisfy equations 4.75.
In addition, for solutions \( \tilde{\beta}_2 \) to equations 4.77, the linear functions

\[
(4.99) \quad U \tilde{\beta}_2,
\]

where \( U = Z'X_2 \) for some \( C(Z) \subseteq C(X_1) \cap C(X_2) \), satisfy

\[
(4.100) \quad U\tilde{\beta}_2 = Z'X_2 \tilde{\beta}_2 = Z'MM_2R_1(M_2R_1)^+y
\]

\[= Z'MR_1(M_2R_1)^+y \]

\[= \phi .\]

In other words, the conditions imposed on the parameters by Wilkinson's algorithm have the form \( \sum n_i a_i = 0 \), not \( \sum a_i = 0 \) as he suggests.

Theorem 4.12 provides the following simpler version of Wilkinson's algorithm. Let \( p'(x) \) denote the reduced minimal polynomial of the matrix \( X_2'X_2 \), then the matrix \( C \) defined by

\[
(4.101) \quad C = q'(X_2'R_1X_2)
\]

with \( q'(x) = [1 - p'(x)]/x \) is a conditional inverse of \( X_2'R_1X_2 \). This would seem simpler for the following reasons:

i) The matrix \( (X_2'X_2)^{-1} \) never need be computed.

ii) The algorithm involves powers of \( X_2'R_1X_2 \) instead of powers of \( R_1M_2R_1 \).

iii) If the distinct nonzero characteristic values of \( X_2'R_1X_2 \) are known, their multiplicities can be determined by calculating \( tr[(X_2'R_1X_2)^i] \).
and then solving the relatively small system of equations

\[(4.102)\]

\[
e_{1\theta_1} + ... + e_{k\theta_k} = \text{tr}[(x'_{21}x_{21})] \\
e_{1\theta_1} + ... + e_{k\theta_k} = \text{tr}[(x'_{21}x_{21})^2] \\
\vdots \\
e_{1\theta_1} + ... + e_{k\theta_k} = \text{tr}[(x'_{21}x_{21})^k]
\]

for the multiplicities \(\theta_1...\theta_k\).

There are of course many other methods for solving the reduced normal equations. Methods involving the reduced minimal polynomial of a matrix are unique. Mann (1960) used this basic method to calculate the orthogonal projection operator on the column space of a matrix A.

F. The Variances of Estimates of Estimable Functions

We can now establish some relationships involving variances of BLUE's of estimable functions. Let \(a'y\) be BLUE for its expectation. Then the variance of \(a'y\) is

\[(4.103)\]

\[\text{Var}(a'y) = \sigma^2 a'a.\]

Since \(a'y\) is BLUE for its expectation, there exists a vector \(\rho\) such that

\[a = x\rho,\]

and
(4.104) \[ \text{Var}(a'y) = \sigma^2 p'X'X p \, . \]

Let \( a'y \) be BLUE for an estimable function of \( \beta_1 \) alone. Then there exists a vector \( \rho_1 \) such that

\[ a = R_1 X_1 \rho_1 \, , \]

and hence

(4.105) \[ \text{Var}(a'y) = \sigma^2 p'X'X R_1 X_1 \rho_1 \, . \]

Similarly, if \( b'y \) is BLUE for an estimable function of \( \beta_2 \) alone, there exists a vector \( \rho_2 \) such that

\[ b = R_2 X_2 \rho_2 \]

and

(4.106) \[ \text{Var}(b'y) = \sigma^2 p'X'X R_2 X_2 \rho_2 \, . \]

Suppose \( c'y \) is BLUE for an estimable function of \( \beta_1 \) and \( \beta_2 \) that cannot be written as a nontrivial sum of estimable functions of \( \beta_1 \) alone and \( \beta_2 \) alone. Then there exists vectors \( \eta_1 \) and \( \eta_2 \) such that

\[ c = X_1 \eta_1 = X_2 \eta_2 \, , \]

and
Next, let \( d'y \) be BLUE for a linear combination of an estimable function in \( \beta_1 \) alone and an estimable function in \( \beta_2 \) alone, i.e. \( d \) can be written as

\[
d = \alpha_1 d_1 + \alpha_2 d_2 ,
\]

where \( d_1 \in C[R_2X_1] \) and \( d_2 \in C[R_1X_2] \). For convenience we will drop the coefficients \( \alpha_1 \) and \( \alpha_2 \) since if \( d_1 \in C[R_2X_1] \) then \( \alpha_1 d \in C[R_2X_1] \) and similarly for \( d_2 \). Then we can write

\[
d = d_1 + d_2 .
\]

Since \( d_1 \in C[R_2X_1] \) and \( d_2 \in C[R_1X_2] \), there exist vectors \( \gamma_1 \) and \( \gamma_2 \) such that

\[
d_1 = R_2X_1\gamma_1 ,
\]

\[
d_2 = R_1X_2\gamma_2 .
\]

This gives

\[
(4.108) \quad \text{Var}(d'y) = \sigma^2 d'd
\]

\[
= \sigma^2(\gamma_1'X_1'R_2 + \gamma_2'X_2'R_1)(R_2X_1\gamma_1 + R_1X_2\gamma_2)
\]

\[
= \sigma^2[\gamma_1'X_1'R_2X_1\gamma_1 + \gamma_2'X_2'R_1X_2\gamma_2 + 2\gamma_1'X_1'R_1X_2\gamma_2]
\]

\[
= \text{Var}(d_1'y) + \text{Var}(d_2'y) + 2 \text{Cov}(d_1'y, d_2'y) .
\]
Now consider a function $\delta'y$ that is BLUE for the sum of an estimable function in $\beta_1$ alone and an estimable function of both $\beta_1$ and $\beta_2$ where the latter estimable function cannot be separated into an estimable function of $\beta_1$ alone and an estimable function of $\beta_2$ alone. That is

$$\delta = \delta_1 + \delta_2$$

where $\delta_1 \in \text{C}(R_2X_1)$ and $\delta_2 \in \text{C}(X_1 \cap X_2)$.

The variance of $\delta'y$ is

$$\text{Var}(\delta'y) = \sigma^2 \delta'\delta$$

$$= \sigma^2 (\delta_1' + \delta_2')(\delta_1 + \delta_2)$$

$$= \sigma^2 (\delta_1'\delta_1 + \delta_2'\delta_2 + 2\delta_1'\delta_2).$$

Since $\delta_1 = R_2X_1\delta_1$ and $\delta_2 = X_2\delta_2$ for some vectors $\delta_1$ and $\delta_2$,

$$\delta_1'\delta_2 = \delta_1'X_1R_2X_2\delta_2$$

$$= \phi.$$ 

This implies that BLUE's of estimable functions of $\beta_1$ alone are uncorrelated with BLUE's of completely confounded functions of $\beta_1$ and $\beta_2$. By a similar development the BLUE's of estimable functions of $\beta_2$ alone are also uncorrelated with confounded functions of $\beta_1$ and $\beta_2$. This gives

$$\text{(4.109)} \quad \text{Var}(\delta'y) = \text{Var}(\delta_1'y) + \text{Var}(\delta_2'y).$$
G. Interfactor Information

Associated with Incomplete Block Designs is the term Interblock Information. The reduced normal equations for treatments eliminating blocks contain intrablock information concerning treatment differences. In some situations it is worthwhile to extract the interblock information. In this section, this concept is generalized to any two-factor situation.

Using the notation and development of Kempthorne (1952), the interblock normal equations can be written as

\[(4.110) \quad B_i = k\mu + \Sigma_{(i)} t_j + k b_i + \Sigma_{(i)} \epsilon_{ij},\]

where \(B_i\) denotes the \(i\)-th block total, \(k\) is the block size, \(b_i\) is the \(i\)-th block effect, and \(\Sigma_{(i)} t_j\) represents the sum of treatment effects for treatments occurring in block \(i\). It is also assumed that the \(b_i\) are uncorrelated with mean zero and constant variance \(\sigma_b^2\). These equations can be rewritten as

\[(4.111) \quad B = Z_1\mu + Z_2 T + \eta,\]

where \(B\) is the \(b \times 1\) vector of block totals, \(Z_1\) is the \(b \times 1\) vector with all elements equal to \(k\), \(Z_2\) is the \(b \times t\) matrix that produces \(\Sigma_{(i)} t_j\), \(T\) is the \(t \times 1\) vector of treatment parameters, and \(\eta\) is a \(b \times 1\) vector of random variables with mean zero and variance-covariance matrix equal to

\[
(4.112) \quad (k \sigma^2 + k^2 \sigma_b^2) I_b.
\]
To change the results of this development to the previous notation, let \( X_1 \) and \( \beta_1 \) correspond to blocks and let \( X_2 \) and \( \beta_2 \) correspond to treatments. Then

\[
B = X_1^T y ,
\]

\[
Z_2^T = X_1^T X_2 \beta_2 ,
\]

and

\[
\eta = X_1^T (X_1 \beta_1 + \epsilon) .
\]

Subtracting the mean, which represents the intersection of \( C(X_1) \) and \( C(X_2) \), gives

\[
X_1^T (I - M)y = X_1^T (I - M)X_2 \beta_2 + \eta ,
\]

where \( \eta \) is a vector of random variables with variance-covariance matrix

\[
\text{Var}(\eta) = X_1^T (I - M)[X_1 V_1 X_1 + \sigma^2 I] (I - M) X_1
\]

and \( V_1 \) is the assumed variance-covariance structure of \( \beta_1 \). The inter-block model 4.114 can be written

\[
z = WB_2 + \eta
\]

where \( \eta \) has variance-covariance structure \( V \) as defined in 4.115. Using the Generalized Normal Equations as defined by Zyskind and Martin (1969),
there exists a class of matrices $V$ that are conditional inverses of $V$ such that BLUE's of estimable functions $\lambda'\beta_2$ are given by $\lambda'\hat{\beta}_2$ where $\hat{\beta}_2$ is any solution to

\[(4.117) \quad W'V^*W\hat{\beta}_2 = W'V^*z\]

and $V^*$ is any matrix in $V$. Clearly if the $\text{rank}(X_1'(I - M)X_2) = \text{rank}(X_1'(I - M))$, then the model in 4.116 is an exact fit for all $y$ and there is no estimate of variance. Therefore, letting $\text{rank}(X_1'(I - M)) = q_{12}$ and $\text{rank}(X_1'(I - M)X_2) = q_0$, the following analysis of variance table can be constructed.

Table 4.6. Interfactor analysis of variance.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments ($\beta_2$)</td>
<td>$q_0$</td>
<td>$\hat{\beta}_2'W'V^*W\hat{\beta}_2$</td>
</tr>
<tr>
<td>Remainder</td>
<td>$q_{12} - q_0$</td>
<td>$\text{difference}$</td>
</tr>
<tr>
<td>Total</td>
<td>$q_{12}$</td>
<td>$z'V^*z$</td>
</tr>
</tbody>
</table>

The development above is completely general and holds for any matrices $X_1$ and $X_2$. To be informative statistically, however, it is necessary to make the assumption that either $\beta_1$ or $\beta_2$ is a vector random variable with mean zero and some variance $V_i$, $i = 1$ or 2. To be useful it is necessary
for \( \text{rank}(X_1'(I - M)) > \text{rank}(X_2'(I - M)X_2) \) or \( \text{rank}(X_2'(I - M)) > \text{rank}(X_1'(I - M)X_1) ) \). Cases when the model 4.114 reduces to an unusable situation are the following:

i) \( X_1'(I - M)X_2 = \phi \)

ii) \( C(X_1) = C(X_2) \).

In either case the matrix \( W \) is null and there is no interfactor information.

It can be seen from the complexity of equations 4.114, 4.116 and 4.117 that for most non-orthogonal designs it may not be worthwhile to extract the interfactor information.
V. THE TWO-WAY FACTORIAL

This chapter and Chapter VI apply the results of Chapter IV to experiments exhibiting the properties of crossing and nesting.

The two-way factorial experiment involves basically two partitions of the model matrix. The mean can be treated as if it were contained in the intersection of $C(X_1)$ and $C(X_2)$. First, the parameters in the two-factor model without the interaction will be examined. Then interaction terms will be introduced and examined.

A. Without Interaction

The simple two-way factorial main-effects model was defined earlier as

\begin{equation}
  y_{ijk} = \mu + f_{i}^{(1)} + f_{j}^{(2)} + \epsilon_{ijk},
\end{equation}

where $i = 1, \ldots, s$, $j = 1, \ldots, t$, $k = 1, \ldots, n_{ij}$. This same model can also be written as

\begin{equation}
  y = 1_n \mu + X_1 \beta_1 + X_2 \beta_2 + \epsilon,
\end{equation}

where $1_n$ is a vector of length $n$ ($n = \Sigma n_{ij}$) with all elements equal to unity, $\beta_1$ is the vector of parameters $f_{i}^{(1)}$, $\beta_2$ is the vector of parameters $f_{j}^{(2)}$, and $X_1$ and $X_2$ are the model matrices corresponding to the two factors.
It was mentioned previously that linear classificatory models are not of full rank. One of the linear relationships among the columns of the model matrix is

\[(5.3)\quad l_n = X_2 l_s = X_3 l_t.\]

For this reason, in this chapter and the next the mean \(\mu\) will be deleted from the model. Therefore, model 5.1 will be written as

\[y = X_1 \beta_1 + X_2 \beta_2 + \epsilon.\]

The effect associated with \(\mu\) will from time to time be discussed.

1. The balanced complete case

It will be informative to first examine briefly the case where \(n_{ij}\) is equal to some constant \(r\). The matrices \(X_1\) and \(X_2\) then have the form

\[(5.4)\quad X_1 = l_r \otimes (l_t \otimes I_s)\]
\[X_2 = l_r \otimes (I_t \otimes l_s),\]

except for a possible reordering of the rows, where the symbol \(\otimes\) denotes the Kronecker product.

The intersection of \(C(X_1)\) and \(C(X_2)\) contains only scalar multiples of the constant vector \(l_n\); therefore,

\[(5.5)\quad M = \frac{1}{n} J_n, \quad I - M = I_n - \frac{1}{n} J_n,\]
where $J_n$ is the $n \times n$ matrix with all elements equal to unity. It can be demonstrated that

\[(5.6) \quad M_1 M_2 = M_2 M_1 = M = \frac{1}{n} J_n.\]

Therefore, the spaces $C(R_2 X_1)$ and $C(R_1 X_2)$ are orthogonal. The rank of the model matrix $(X_1, X_2)$ is $s + t - 1$, $\text{rank}(M) = 1$, $\text{rank}(R_2 X_1) = s - 1$, and $\text{rank}(R_1 X_2) = t - 1$. The form of the analysis of variance table is well known and is displayed in Table 5.1.

Table 5.1. The balanced, complete two-way factorial.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounded effects $(\mu)$</td>
<td>1</td>
<td>$y' M y$</td>
</tr>
<tr>
<td>First factor $(\beta_1)$</td>
<td>$s - 1$</td>
<td>$y'(M_1 - M) y$</td>
</tr>
<tr>
<td>Second factor $(\beta_2)$</td>
<td>$t - 1$</td>
<td>$y'(M_2 - M) y$</td>
</tr>
<tr>
<td>Error</td>
<td>$rst - t - s + 1$</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>$rst$</td>
<td>$y'y$</td>
</tr>
</tbody>
</table>

All contrasts of the form $f_i^{(1)} - f_j^{(1)}$ and $f_i^{(2)} - f_j^{(2)}$ are estimable and their BLUE's are given by corresponding differences of treatment means.
2. The general case

The case when the numbers \( n_{ij} \) are unequal but proportionate also admits a decomposition similar to that in Table 5.1. A discussion of proportionate subclass numbers can be found in Kempthorne (1952) and will not be covered here. Instead, we proceed directly to the general case where \( n_{ij} \geq 0 \).

When \( n_{ij} > 0 \) for all \( i \) and \( j \), the model is of maximal rank and the only vectors in the intersection of \( C(X_1) \) and \( C(X_2) \) are scalar multiples of the constant vector \( 1_n \). The orthogonal projection operator \( M \) has the form

\[
(5.7) \quad M = \frac{1}{n} J_n,
\]

where \( n = \sum n_{ij} \). The rank of \( R_{1}X_1 \) is \( s - 1 \) and rank of \( R_{1}X_2 \) is \( t - 1 \). However, the spaces \( R_{2}X_1 \) and \( R_{1}X_2 \) are no longer orthogonal. The analysis of variance can be displayed as in Table 5.2, which is a special case of Table 4.1.

If \( n_{ij} = 0 \) for some values of \( i \) and \( j \), the intersection of \( C(X_1) \) and \( C(X_2) \) may contain more than multiples of the constant vector. Then the analysis of variance table would be identical to Table 4.1. Note that line (1) of Table 4.1 can be partitioned further to remove the single degree of freedom for the mean.

In some special cases, the three spaces \( C(M), C(R_{2}X_1), \) and \( C(R_{1}X_2) \) may be orthogonal, even for unequal numbers and missing cells. For these designs the analysis of variance has the form illustrated in Table 5.3.
### Table 5.2. Two-way factorial with unequal numbers but no missing cells.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of Squares</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounded effects (µ)</td>
<td>y'M y</td>
<td>1</td>
<td>y'M y</td>
<td>Confounded effects (µ)</td>
</tr>
<tr>
<td>F\textsubscript{1} eliminating F\textsubscript{2}</td>
<td>β\textsuperscript{1}'X\textsuperscript{1}R\textsuperscript{2}X\textsuperscript{2}β\textsuperscript{2} \textsuperscript{1}</td>
<td>s - 1</td>
<td>β\textsuperscript{1}'(I-M)X\textsuperscript{1}β\textsuperscript{2} \textsuperscript{1}</td>
<td>Remainder</td>
</tr>
<tr>
<td>Remainder</td>
<td>β\textsuperscript{2}'X\textsuperscript{1}(I-M)X\textsuperscript{2}β\textsuperscript{2} \textsuperscript{2}</td>
<td>t - 1</td>
<td>β\textsuperscript{2}'X\textsuperscript{1}R\textsuperscript{2}X\textsuperscript{2}β\textsuperscript{2} \textsuperscript{2}</td>
<td>F\textsubscript{2} eliminating F\textsubscript{1}</td>
</tr>
<tr>
<td>Error</td>
<td>difference</td>
<td>n - s - t + 1</td>
<td>difference</td>
<td>Error</td>
</tr>
</tbody>
</table>

| Total                           | y'y             | n                  |                 |                    |
Table 5.3. Special case of an orthogonal decomposition.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely confounded effects</td>
<td>rank(M)</td>
<td>y'M y</td>
</tr>
<tr>
<td>Factor F_1 (alone)</td>
<td>rank((I-M)X_1)</td>
<td>y'(M_1-M)y</td>
</tr>
<tr>
<td>Factor F_2 (alone)</td>
<td>rank((I-M)X_2)</td>
<td>y'(M_2-M)y</td>
</tr>
<tr>
<td>Error</td>
<td>n - rank(X_1,X_2)</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>n</td>
<td>y'y</td>
</tr>
</tbody>
</table>

The experiment introduced in Chapter II to illustrate the algorithm due to Wilkinson admits the above orthogonal decomposition. This experiment will be used here to illustrate the results of Chapter IV.

Example 5.1 Consider one replicate of the two-factor additive experiment from Chapter II. The basic design configuration is illustrated in Figure 5.1. Asterisks denote observed treatment combinations.
The model can be written as

\[(5.8) \quad y_{ij} = \mu + f_i^{(1)} + f_j^{(2)} + \varepsilon_{ij},\]

where \((i,j) = (1,1), (1,3), (2,2), (2,4), (3,1), (3,3), (4,2)\) and \((4,4)\).

The matrix model equation can be written (deleting the mean as was done previously) as

\[(5.9) \quad y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon,\]

or more specifically, as
A basis for $C(X_1) \cap C(X_2)$ can be made up of the two columns of the matrix,

\[
(5.11) \begin{pmatrix}
1 & 1 \\
1 & 1 \\
1 & 0 \\
1 & 0 \\
1 & 1 \\
1 & 1 \\
1 & 0 \\
1 & 0
\end{pmatrix}
\]
It can be seen that the first column of the matrix in 5.11 can be formed by summing the columns of \( X_1 \) or \( X_2 \). The second column in 5.11 can be formed from the sum of the first and third columns of either \( X_1 \) or \( X_2 \). Any other vectors in \( C(X_1) \cap C(X_2) \) can be generated by taking a linear combination of the vectors in 5.11.

The orthogonal projection operator \( M \) can be calculated and shown to be

\[
M = \frac{1}{4}
\begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1
\end{bmatrix}
\]

The matrices \( R^{X_1} \) and \( R^{X_2} \) can be written

\[
R^{X_1} = \frac{1}{2}
\begin{bmatrix}
1 & 0 & -1 & 0 \\
1 & 0 & -1 & 0 \\
0 & 1 & 0 & -1 \\
0 & 1 & 0 & -1 \\
-1 & 0 & 1 & 0 \\
-1 & 0 & 1 & 0 \\
0 & -1 & 0 & 1 \\
0 & -1 & 0 & 1
\end{bmatrix},
R^{X_2} = \frac{1}{2}
\begin{bmatrix}
1 & 0 & -1 & 0 \\
-1 & 0 & 1 & 0 \\
0 & 1 & 0 & -1 \\
0 & -1 & 0 & 1 \\
1 & 0 & -1 & 0 \\
-1 & 0 & 1 & 0 \\
0 & 1 & 0 & -1 \\
0 & -1 & 0 & 1
\end{bmatrix}
\]
It can easily be seen by examining these three matrices that the following properties hold:

i) \( r(R_1^Xg) = \text{dim}[C(R_1^Xg)] = 2 \)

ii) \( r(R_2^Xg) = \text{dim}[C(R_2^Xg)] = 2 \)

iii) \( r(M) = \text{dim}[C(X_1) \cap C(X_2)] = 2 \)

iv) \( X_1^R_2^Xg \) = \( \phi \)

From these properties, it follows that:

i) there are two linearly independent estimable functions of \( \beta_1 \) alone,

ii) there are two linearly independent estimable functions of \( \beta_2 \) alone,

iii) there is one completely confounded estimable function of \( \beta_1 \) and \( \beta_2 \) (in addition to the mean), and

iv) the decomposition of the sum of squares is an orthogonal decomposition which implies that estimates of estimable functions are uncorrelated.

The analysis of variance can be represented symbolically as in Table 5.4.
Table 5.4. Symbolic representation of the analysis.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely confounded effects</td>
<td>2</td>
<td>$y'M\ y$</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>$\frac{1}{8}y'J_gy$</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>$y'(M - \frac{1}{8}J_g)y$</td>
</tr>
<tr>
<td>Factor $F_1$ alone</td>
<td>2</td>
<td>$y'(M_1-M)y$</td>
</tr>
<tr>
<td>Factor $F_2$ alone</td>
<td>2</td>
<td>$y'(M_2-M)y$</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>$y'y$</td>
</tr>
</tbody>
</table>

The specific sum of squares can be calculated using $M$ as defined in 5.12, $M_1 = X_1(X_1'X_1)^{-1}X_1$, and $M_2 = X_2(X_2'X_2)^{-1}X_2$.

Estimable functions of $\beta_1$ alone and $\beta_2$ alone can be written as $\lambda_1'\beta_1$ and $\lambda_2'\beta_2$ where $\lambda_1 \in R(R_2X_1)$ and $\lambda_2 \in R(R_1X_2)$. By examining the rows of $R_2X_1$ and $R_1X_2$ in expression 5.13, it can easily be seen that a set of linearly independent estimable functions is:
The completely confounded effects are determined by \( R(\mathbf{X}_1, \mathbf{X}_2) \).

The matrix \((\mathbf{X}_1, \mathbf{X}_2)\) is equal to

\[
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}
\]

Inspection of linear combinations of the rows of these matrices yields the following independent set of estimable functions,

\[
(\mathbf{5.16}) \quad f_1^{(1)} + f_2^{(1)} + f_3^{(1)} + f_4^{(1)} + f_1^{(2)} + f_2^{(2)} + f_3^{(2)} + f_4^{(2)}
\]

\[
f_1^{(1)} + f_3^{(1)} + f_1 + f_3
\]
Table 5.5. Expected mean squares for example 5.1.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random Effects Model</strong> (E(β_i) = φ, Var(β_i) = σ_i^2I)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete confounding</td>
<td>2</td>
<td>σ^2 + 2σ_1^2 + 2σ_2^2</td>
</tr>
<tr>
<td>Factor F_1 alone</td>
<td>2</td>
<td>σ^2 + 2σ_1^2</td>
</tr>
<tr>
<td>Factor F_2 alone</td>
<td>2</td>
<td>σ^2 + 2σ_2^2</td>
</tr>
<tr>
<td><strong>Fixed Effects Model</strong> (β_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4}))**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete confounding</td>
<td>2</td>
<td>σ^2 + 2[(b_{11} + b_{13})^2 + (b_{12} + b_{14})^2]</td>
</tr>
<tr>
<td>Factor F_1 alone</td>
<td>2</td>
<td>σ^2 + 1/2[(b_{11} - b_{13})^2 + (b_{12} - b_{14})^2]</td>
</tr>
<tr>
<td>Factor F_2 alone</td>
<td>2</td>
<td>σ^2 + 1/2[(b_{21} - b_{23})^2 + (b_{22} - b_{24})^2]</td>
</tr>
<tr>
<td><strong>Mixed Effects Model</strong> (A-fixed, B-random)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete confounding</td>
<td>2</td>
<td>σ^2 + 1/2[(b_{11} - b_{13})^2 + (b_{12} - b_{14})^2] + 2σ_2^2</td>
</tr>
<tr>
<td>Factor F_1 alone</td>
<td>2</td>
<td>σ^2 + 1/2[(b_{11} - b_{13})^2 + (b_{12} - b_{14})^2]</td>
</tr>
<tr>
<td>Factor F_2 alone</td>
<td>2</td>
<td>σ^2 + 2σ_2^2</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>σ^2</td>
</tr>
</tbody>
</table>
The first of these functions can be thought of as the overall mean \( \mu \). The second is sum of a function of \( \beta_1 \) and a function of \( \beta_2 \). If this function is to be treated as a function of \( f_1 \), for example, then the experimenter must have evidence that \( f_1^{(2)} + f_3^{(2)} = 0 \).

The expected mean squares for the random effects model, fixed effects model, and mixed effects model can be calculated and are displayed in Table 5.5.

B. With Interaction

The two-way factorial experiment with interaction is a special case of the situation where part of the model matrix \( X \) is completely contained in the column space generated by the rest of the model matrix. The general case is discussed in the next chapter.

The model for a two-factor experiment with interaction was defined earlier as

\[
Y_{ijk} = \mu + f_i^{(1)} + f_j^{(2)} + f_{ij}^{(1,2)} + \varepsilon_{ijk},
\]

with \( i = 1, \ldots, s \), \( j = 1, \ldots, t \), and \( k = 1, \ldots, n_{ij} \). In matrix notation the model can be written as

\[
y = X_1 \beta_1 + X_2 \beta_2 + X_3 \beta_3 + \varepsilon,
\]

where the mean effect \( \mu \) has been deleted, the matrices \( X_1, \beta_1, X_2 \) and \( \beta_2 \) are identical to those defined in the previous section, \( \beta_3 \)
corresponds to the interaction parameters and \( X_3 \) is the corresponding model matrix.

1. **The balanced complete case**

   As in the previous section, the case \( n_{ij} = r \) (a constant) will be examined first. The matrices \( X_1 \) and \( X_2 \) have the form defined in 5.4. The matrix \( X_3 \) can be written as

   \[
   (5.19) \quad X_3 = l_r \otimes I_{st},
   \]

   where \( \otimes \) denotes the Kronecker product.

   Since the matrices \( I_t \otimes I_s \) and \( I_t \otimes I_s \) can clearly be written as linear combinations of columns of \( I_{st} \), it follows that \( C(X_1, X_2) \subseteq C(X_3) \). The reduced normal equations for \( \beta_3 \) eliminating \( \beta_1 \) and \( \beta_2 \) are

   \[
   (5.20) \quad X_3^R X_3 \beta_3 = X_3^R y,
   \]

   where \( R \) is the orthogonal projection operator on \( C(X_1, X_2) \). From Theorem 3.1 these equations can be solved by imposing the conditions

   \[
   (5.21) \quad (X_1, X_2)^T X_3 \beta_3 = \phi
   \]

   on the model

   \[
   (5.22) \quad y = X_3 \beta_3 + \eta.
   \]
The conditions 5.21 are estimable with model 5.22 and for the case 
\( n_{ij} = r \) have the form 

\[
\sum_i f_{ij}^{(1,2)} = 0 \text{ for all } j,
\]

\[
\sum_j f_{ij}^{(1,2)} = 0 \text{ for all } i.
\]

These are the standard conditions usually imposed to obtain unique solutions to the normal equations.

Estimates of main-effects are defined as 

\[
\hat{f}_i^{(1)} = y_{i..} - \bar{y}_{..},
\]

where the \((\cdot)\) signifies calculating the mean over that specific subscript. Treatment differences \( f_i^{(1)} - f_m^{(1)} \) are estimated by differences in the corresponding main-effects defined by 5.24. This gives 

\[
\hat{f}_i^{(1)} - \hat{f}_m^{(1)} = y_{i..} - y_{m..}.
\]

This expression has expectation 

\[
\mathbb{E}[y_{i..} - y_{m..}] = f_i^{(1)} - f_m^{(1)} + \frac{1}{rt} \left( \sum_j f_{ij}^{(1,2)} - \sum_{ml} f_{ml}^{(1,2)} \right).
\]

From conditions 5.23 we say that the summations in 5.26 are zero; and therefore, 5.25 is at least a conditionally unbiased estimate of 
\( f_i^{(1)} - f_m^{(1)} \).
It is of interest to note that when interaction terms are included in the model there are no unbiased estimates of main-effect differences. Every estimable function involving main-effects also involves interaction parameters. This causes serious inferential problems when interaction effects are nonzero.

The analysis of variance for the balanced complete case is displayed in Table 5.6.

Table 5.6. The balanced complete case with interaction.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confounded effects (μ)</td>
<td>1</td>
<td>y'M y</td>
</tr>
<tr>
<td>Factor F_1</td>
<td>s - 1</td>
<td>y'(M_1-M)y</td>
</tr>
<tr>
<td>Factor F_2</td>
<td>t - 1</td>
<td>y'(M_2-M)y</td>
</tr>
<tr>
<td>Interaction</td>
<td>(s - 1)(t - 1)</td>
<td>y'(M_3-M_1-M_2+M)y</td>
</tr>
<tr>
<td>Error</td>
<td>st(r - 1)</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>rst</td>
<td>y'y</td>
</tr>
</tbody>
</table>

2. The general case

When the subclass numbers n_{ij} are not constant, in fact some may be zero, the problems associated with estimation become more complex. The basic reason for the complexity is that the conditions imposed on the parameters to obtain unique solutions to the normal equations appear in the
definitions of main-effects. In other words, with one set of conditions, differences of treatment means are at least conditionally unbiased estimates for the same differences of treatment effects. However, with another set of conditions these same differences of treatment means are not even conditionally unbiased estimates of treatment effect differences.

The conditions imposed on the interaction parameters are arbitrary except that they are nonestimable and consistent. A general form for them would be

\[(5.27) \quad \sum_{i} u_i f_{ij}^{(1,2)} = 0 \text{ for all } j, \]
\[(5.28) \quad \sum_{j} v_j f_{ij}^{(1,2)} = 0 \text{ for all } i, \]

where \( \sum_{i} u_i = \sum_{j} v_j = 1 \). Two distinct sets of \( u_i, v_j \) will be examined:

\[(5.28a) \quad u_1 = \sum_{ij} n_{ij} / \sum_{ij} n_{ij}, \quad v_1 = \sum_{ij} n_{ij} / \sum_{ij} n_{ij} \]
\[(5.28b) \quad u_2 = \ldots = u_s, \quad v_2 = \ldots = v_t. \]

In the general case, \( n_{ij} \geq 0 \), the properties \( C(X_1) \subseteq C(X_3) \) and \( C(X_2) \subseteq C(X_3) \) still hold. The reduced normal equations for \( \beta_3 \) eliminating \( \beta_1 \) and \( \beta_2 \) have the form

\[(5.29) \quad X_3' R X_3 \beta_3 = X_3' R y, \]
where $R$ is the orthogonal projection operator on $C_1(x_1, x_2)$. From Theorem 3.1 it follows that solutions to equations 5.29 can be obtained by imposing the estimable conditions

\[(5.30) \quad (x_1, x_2)'x_3\hat{\beta}_3 = 0\]

on the model

\[(5.31) \quad y = x_3\hat{\beta}_3 + \eta .\]

The resulting normal equations are

\[(5.32) \quad X_3'x_3\hat{\beta}_3 = x_3'Ry .\]

The solutions $\hat{\beta}_1$ and $\hat{\beta}_2$ are obtained by solving

\[\begin{pmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} (y - x_3\hat{\beta}_3) .\]

From conditions 5.30 it follows that the solutions $\hat{\beta}_1$ and $\hat{\beta}_2$ are the solutions obtained by ignoring the interaction parameters. The above procedure is straightforward, computationally. However, it appears as though the data are dictating the conditions to be imposed on the parameters to obtain unique solutions. From an inferential standpoint, this is a problem.
The main-effects are defined to be the vectors $\hat{\beta}_1$ and $\hat{\beta}_2$. Estimable functions involving $\beta_1$ and $\beta_2$ have the form

$$
\lambda_1^T \beta_1 + \lambda_2^T \beta_2 + \lambda_3^T \beta_3,
$$

with $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}(X_1, X_2, MX_3)$ where $M$ is the orthogonal projection operator on $C(X_1, X_2)$. Vectors in $\mathbb{R}(X_1, X_2, MX_3)$ of the form

$$
\rho'(X_1, X_2, MX_3)
$$

where $\rho$ is any vector in $C(R_2 X_1)$ produce estimable functions of the form

$$
\rho' X_1 \beta_1 + \rho' MX_3 \beta_3.
$$

Since the expression $\rho' MX_3 \beta_3$ is conditionally zero (conditional upon 5.30) for any $\rho$, it follows that functions of the form $\lambda_1^T \beta_1$ are conditionally estimable for any $\lambda_1 \in \mathbb{R}(X_1 R_2 X_1)$. From the definition of $M$ it follows that $X_1^T R_2 M = X_1^T R_2 M$; therefore, for any vector $a \in C(R_2 X_1)$ the function $a'y$ is a conditional BLUE of a function $\lambda_1^T \beta_1$. Similar results can be obtained for functions of $\beta_2$. It also follows that complete confounding between functions of $\beta_1$ and functions of $\beta_2$ can be defined within the framework of conditions 5.30. It should be noted that expect for very special cases, differences of treatment means are not conditional BLUE's of the same differences of treatment effects. Conditions 5.30 are, except for a scalar multiplier, conditions (1) of 5.28.
The conditions (2) of 5.28 are the same conditions that are imposed when \( n_{ij} = r \) (a constant) for all \( i \) and \( j \). These conditions have an advantage over the others in that they are independent of the data.

When there are no missing cells and the mean \( \mu \) is included in the model, the estimates are usually written as

\[
\hat{\mu} = y_{...},
\]

\[
\hat{\beta}_1 = \begin{pmatrix} y_{1...} - y_{...} \\ \vdots \\ y_{s...} - y_{...} \end{pmatrix},
\]

\[
\hat{\beta}_2 = \begin{pmatrix} y_{1.} - y_{...} \\ \vdots \\ y_{t.} - y_{...} \end{pmatrix},
\]

\[
\hat{\beta}_3 = \begin{pmatrix} y_{11} - y_{1...} - y_{1.} + y_{...} \\ \vdots \\ y_{st} - y_{s...} - y_{t.} + y_{...} \end{pmatrix},
\]

where \( y_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ij} \), \( y_{i...} = \sum_{j} y_{ij} / t \), \( y_{.j} = \sum_{i} y_{ij} / s \), and

\( y_{...} = \frac{1}{st} \sum_{ij} y_{ij} \). Since the function
can be written as $a'y$ with $a \in C(X_1,X_2,X_3)$, it follows that (5.42) is BLUE of its expectation which is

$$E[y_{1..} - y_.] = f_i^{(1)} - f_i^{(1)} + \frac{1}{t} \sum_j f_{ij}^{(1,2)} + \frac{1}{s} \sum_i f_{ij}^{(1,2)}.$$ 

From conditions (2) of 5.28 it follows that the expressions involving the $f_{ij}^{(1,2)}$ should be zero. Therefore, differences of treatment means are conditional BLUE's (conditional on (2) of 5.28) of the same differences of treatment effects.

When at least one cell is missing, say the 1,1-th cell, it becomes somewhat obscure how the condition

$$\sum_j f_{ij}^{(1,2)} = 0$$

is to be imposed on the solutions.

However, regardless of which conditions are imposed on the solutions to the normal equations

$$(5.44) \begin{bmatrix} x_1'x_1 & x_1'x_2 & x_1'x_3 \\ x_2'x_1 & x_2'x_2 & x_2'x_3 \\ x_3'x_1 & x_3'x_2 & x_3'x_3 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} x_1'y_1 \\ x_2'y_1 \\ x_3'y_1 \end{bmatrix},$$

the decomposition of $\hat{y}'\hat{y}$ into sum of squares as in Table 5.7 is unique.
Table 5.7. The two-way factorial experiment with interaction.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Quadratic Form*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>$y'M_0y$</td>
</tr>
<tr>
<td>$F_1$ and $F_2$ confounded</td>
<td>$q_{12}-1$</td>
<td>$y'(M-M_0)y$</td>
</tr>
<tr>
<td>$F_1$ alone</td>
<td>$q_1-q_{12}$</td>
<td>$\hat{\beta}_1'X_1'X_1\hat{\beta}_1$</td>
</tr>
<tr>
<td>$F_2$ alone</td>
<td>$q_2-q_{12}$</td>
<td>$\hat{\beta}_2'X_2'X_2\hat{\beta}_2$</td>
</tr>
<tr>
<td>Remainder</td>
<td>---</td>
<td>$S^*$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$q-q_1-q_2+q_{12}$</td>
<td>$\hat{\beta}_3'X_3'X_3\hat{\beta}_3$</td>
</tr>
<tr>
<td>Error</td>
<td>$n-q$</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>$n$</td>
<td>$y'y$</td>
</tr>
</tbody>
</table>

$$S^* = \begin{pmatrix} \hat{\beta}_1' \\ \hat{\beta}_2' \end{pmatrix} \begin{pmatrix} X_1'(M_2-M)X_1 & X_1'(M-M_2)X_2 \\ X_2'(M-M_2)X_1 & X_2'(M_1-M)X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

*All quadratic forms except that for "Remainder" are sums of squares.

To determine appropriate ratios of mean squares for testing certain hypotheses we next turn our attention to the calculation of expected mean squares. The procedure developed in Chapter IV will be used here to examine fixed effects, random effects, and mixed effects models.

The fixed effects model hypothesizes that the parameters $\beta_1$, $\beta_2$, and $\beta_3$ are vectors of fixed constants $a$, $b$ and $g$. Under this assumption $yy'$ has expectation
The expected mean squares for the quantities in Table 5.11 can be calculated and are presented in Table 5.8.

For the random effects hypothesis, the parameters $\beta_1$, $\beta_2$, and $\beta_3$ are assumed to be independently distributed random variables with means zero and variances $\sigma^2_{1s}$, $\sigma^2_{2t}$, and $\sigma^2_{12}$, respectively. Under these assumptions the expectation of $yy'$ is

$$E[yy'] = \sigma^2 I$$

$$+ \sigma^2_{11}X_1X_1' + \sigma^2_{22}X_2X_2' + \sigma^2_{12}X_3X_3' .$$

Applying this result to the expressions in Table 5.7 gives the expected mean squares displayed in Table 5.9.

The mixed effects hypothesis has been a subject of discussion for some time. We will present the "usual" assumptions as well as a more appropriate alternative.

The "usual" assumptions are as follows: assume $\beta_1$ is a vector of fixed constants $a$, $\beta_2$ is a vector of random variables with mean zero and variance $\sigma^2_{2t}$, and $\beta_3$ is a vector of random variables with mean zero and variance $\sigma^2_{12}$. With these assumptions the quantity $yy'$ has expectation

$$E[yy'] = \sigma^2 I$$

$$+ \sigma^2_{11}X_1X_1' + \sigma^2_{22}X_2X_2' + \sigma^2_{12}X_3X_3' .$$
Table 5.8. Expected mean squares for the fixed effects two-way factorial with interaction.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$ and $F_2$ confounded</td>
<td>$q_{12} - 1$</td>
<td>$\sigma^2 + (x_1a + x_2b + x_3g)'(M-M_0)(x_1a + x_2b + x_3g)/(q_{12} - 1)$</td>
</tr>
<tr>
<td>$F_1$ alone</td>
<td>$q_1 - q_{12}$</td>
<td>$\sigma^2 + (x_1a + x_3g)'R_1X_1(X_1'X_1)^{-1}X_2'X_1(x_1a + x_3g)/(q_1 - q_{12})$</td>
</tr>
<tr>
<td>$F_2$ alone</td>
<td>$q_2 - q_{12}$</td>
<td>$\sigma^2 + (x_2b + x_3g)'R_2X_2(X_2'X_2)^{-1}X_1'X_2(x_2b + x_3g)/(q_2 - q_{12})$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$q - q_1 - q_2 + q_{12}$</td>
<td>$\sigma^2 + (x_3g)'R(x_3g)/(q - q_1 - q_2 + q_{12})$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - q$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
Table 5.9. Expected mean squares for the random effects two-way factorial with interaction.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$ and $F_2$ confounded</td>
<td>$q_{12}-1$</td>
<td>$\sigma^2 + \frac{\sigma_1^2 \text{tr}(X_1'M_0X_1)}{(q_{12}-1)}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ $\frac{\sigma_2^2 \text{tr}(X_2'M_0X_2)}{(q_{12}-1)}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ $\frac{\sigma_{12}^2 \text{tr}(X_3'M_0X_3)}{q_{12}-1}$</td>
</tr>
<tr>
<td>$F_1$ alone</td>
<td>$q_1 - q_{12}$</td>
<td>$\sigma^2 + \frac{\sigma_1^2 \text{tr}(X_1'R_1X_1)}{(q_1-q_{12})}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ $\frac{\sigma_2^2 \text{tr}(X_1'R_2X_1X_2)}{(q_{12}-q_1)}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ $\frac{\sigma_{12}^2 \text{tr}(X_3'R_2X_1X_2X_3)}{(q_{12}-q_1)}$</td>
</tr>
<tr>
<td>$F_2$ alone</td>
<td>$q_2 - q_{12}$</td>
<td>$\sigma^2 + \frac{\sigma_1^2 \text{tr}(X_1'R_1X_2)}{(q_2-q_{12})}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ $\frac{\sigma_2^2 \text{tr}(X_3'R_1X_2X_3)}{(q_{12}-q_2)}$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$q - q_1 - q_2 + q_{12}$</td>
<td>$\sigma^2 + \frac{\sigma_{12}^2 \text{tr}(X_1'R_1X_3)}{(q-q_1-q_2+q_{12})}$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - q$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
Using this quantity, the expected mean squares for the expressions in Table 5.7 can be calculated and are given Table 5.10.

The more appropriate alternative to the "usual" mixed effects model assumptions is based on the fact that even though the vector $\beta_3$ is a vector of random variables, it satisfies some constraints, namely,

\begin{equation}
\Sigma_{i=1}^{t} \sum_{j=1}^{(1,2)} f_{ij} = 0 \quad \text{for all } i .
\end{equation}

For this reason, the vector $\beta_3$ is assumed to have a variance-covariance matrix

\begin{equation}
V = \begin{pmatrix}
V_1 & \phi \\
\phi & V_2 \\
\end{pmatrix},
\end{equation}

where each $V_i$ has the form

\begin{equation}
V_i = \sigma^2_{12} (I_t - \frac{1}{t} J_t) .
\end{equation}

The matrix $X_3$ can be partitioned correspondingly into
The expected value of $yy'$ can then be written as

$$E[yy'] = \sigma^2 I + (x_1 a)(x_1 a)'$$

$$+ \sigma^2 X_2 X_2' + \sigma^2 \sum_{i=1}^{s} X_{3i} J_{3i} X_{3i}' .$$

Substituting (5.50) in the above expression gives

$$E[yy'] = \sigma^2 I + (x_1 a)(x_1 a)' + \sigma^2 X_2 X_2'$$

$$+ \sigma^2 \sum_{i=1}^{s} X_{3i} X_{3i}' - \frac{1}{t} \sum_{i=1}^{s} X_{3i} X_{3i}' .$$

Let $J_t = l_{11}^t$ and define

$$M_{12} = R_1 X_2 (X_2' R_2 X_2)' X_2^t R_1$$

$$M_{21} = R_2 X_1 (X_1' R_1 X_1)' X_1^t R_2 ,$$

then the expected mean squares for the quantities in Table 5.7 can be written as in Table 5.11.
Table 5.10. Expected mean square for the mixed effects two-way factor with interaction (Case I).

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>( F_1 ) and ( F_2 ) confounded</td>
<td>( q_{12} )</td>
<td>( \sigma^2 + (X_1a)'(M-M_0)(X_1a)/q_{12} + \sigma^2 tr(X_2'(M-M_0)X_2)/q_{12} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ ( \sigma^2_{12} tr(X_3'(M-M_0)X_3)/q_{12} )</td>
</tr>
<tr>
<td>( F_1 ) alone</td>
<td>( q_1 - q_{12} )</td>
<td>( \sigma^2 + a'R_1X_1a/(q_1 - q_{12}) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ ( \sigma^2_{12} tr(X_3'[R_1X_1(X_1'R_1X_1)'X_1'R_1]X_3)/(q_1 - q_{12}) )</td>
</tr>
<tr>
<td>( F_2 ) alone</td>
<td>( q_2 - q_{12} )</td>
<td>( \sigma^2 + \sigma^2_{12} tr(X_2'R_2X_2)/(q_2 - q_{12}) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ ( \sigma^2_{12} tr(X_3'[R_2X_2(X_2'R_2X_2)'X_2'R_2]X_3)/(q_2 - q_{12}) )</td>
</tr>
<tr>
<td>Interaction</td>
<td>( q - q_1 - q_2 + q_{12} )</td>
<td>( \sigma^2 + \sigma^2_{12} tr(X_3'R_3X_3)/(q - q_1 - q_2 + q_{12}) )</td>
</tr>
<tr>
<td>Error</td>
<td>( n - q )</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>
### Table 5.11. Expected mean squares for the mixed effects two-way factorial with interaction (Case II).

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F₁ and F₂ confounded</strong></td>
<td>q₁₂</td>
<td>$\sigma^2 + (X_1 a)'(M-M_0)(X_1 a)'q_{12} + \sigma^2 r(X_2(M-M_0)X_2)/q_{12}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 + \sigma^2 \left[ \frac{1}{t} \sum_{i=1}^{s} tr(X_{31i}'(M-M_0)X_{31i}) - \frac{1}{t} \sum_{i=1}^{s} 1'X_{31i}'(M-M_0)X_{31i}1_t / t \right] / q_{12}$</td>
</tr>
<tr>
<td><strong>F₁ alone</strong></td>
<td>q₁ - q₁₂</td>
<td>$\sigma^2 + a'X_1'RX_1a/(q_1 - q_{12})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 + \sigma^2 \left[ \frac{1}{t} \sum_{i=1}^{s} tr(X_{31i}'M_0X_{31i}) - \frac{1}{t} \sum_{i=1}^{s} 1'X_{31i}'M_0X_{31i}1_t / t \right] / (q_1 - q_{12})$</td>
</tr>
<tr>
<td><strong>F₂ alone</strong></td>
<td>q₂ - q₁₂</td>
<td>$\sigma^2 + \sigma^2 r(X_2'RX_2)/(q_2 - q_{12})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 + \sigma^2 \left[ \frac{1}{t} \sum_{i=1}^{s} tr(X_{31i}'M_2X_{31i}) - \frac{1}{t} \sum_{i=1}^{s} 1'X_{31i}'M_2X_{31i}1_t / t \right] / (q_2 - q_{12})$</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>q₁ - q₁ - q₂ + q₁₂</td>
<td>$\sigma^2 + \sigma^2 \left[ \frac{1}{t} \sum_{i=1}^{s} tr(X_{31i}'RX_{31i}) - \frac{1}{t} \sum_{i=1}^{s} 1'X_{31i}'RX_{31i}1_t / t \right] / (q_1 - q_2 + q_{12})$</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>n - q</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
VI. NESTED FACTORS

In this chapter the linear model

\begin{equation}
\begin{split}
y_{ijk} &= \mu + f_{i}^{(1)} + f_{ij}^{(2)} + \varepsilon_{ijk},
\end{split}
\end{equation}

where \(i = 1, 2, \ldots, s\), \(j = 1, 2, \ldots, t\), \(k = 1, 2, \ldots, n_{ij}\), will be examined using the results and notation of previous chapters. The matrix representation of 6.1 can be written

\begin{equation}
\begin{split}
y &= 1_{n} \mu + X_{1} \beta_{1} + X_{2} \beta_{2} + \varepsilon,
\end{split}
\end{equation}

where \(n = \sum n_{ij}\), \(\beta_{1}\) corresponds to the vector of parameters \((f_{i}^{(1)}, \ldots, f_{i}^{(1)})\), \(\beta_{2}\) corresponds to the vector of parameters \((f_{11}^{(2)}, \ldots, f_{st}^{(2)})\), and \(X_{1}\) and \(X_{2}\) are the corresponding model matrices. Since the vector \(1_{n}\) can be written

\begin{equation}
\begin{split}
1_{n} = X_{1} 1_{s} = X_{2} 1_{st},
\end{split}
\end{equation}

the mean \(\mu\) will be deleted from the model. Its contribution, however, will be computed when it is appropriate. Therefore, the basic model under discussion is

\begin{equation}
\begin{split}
y &= X_{1} \beta_{1} + X_{2} \beta_{2} + \varepsilon,
\end{split}
\end{equation}

where \(C(X_{1}) \subseteq C(X_{2})\).
A. Estimation

1. The balanced complete case

The case when \( n_{ij} = r \) (a constant) for all \( i \) and \( j \) will be discussed briefly to introduce the simple nested model. In this case the model matrices \( X_1 \) and \( X_2 \) have the form

\[
X_1 = 1_r \otimes (1_t \otimes I_s)
\]

and

\[
X_2 = 1_r \otimes I_{st}
\]

except for a possible reordering of the rows. To obtain a set of solutions to the normal equations the nonestimable conditions

\[
\sum_{j=1}^{t} f_{ij}^{(2)} = 0 \quad \text{for all } i
\]

are usually imposed on the parameters. These conditions can be written alternatively as

\[
X_1'X_2\beta_2 = \phi
\]

The reduced normal equations for \( \beta_2 \) eliminating \( \beta_1 \) are

\[
X_2'X_2\beta_2 = X_2'\mathbf{y}
\]
Since \( C(X_1) \subseteq C(X_2) \), Theorem 3.1 can be used to derive solutions to equations 6.9. That is, solutions to equations 6.9 can be obtained by imposing the estimable conditions

\[(6.10) \quad X'_1X_2\beta_2 = \phi\]

on the model

\[(6.11) \quad y = X_2\beta_2 + \eta.\]

The appropriate normal equations are

\[(6.12) \quad X'_2X_2\hat{\beta}_2 = X'_2R_2y.\]

The corresponding solutions \( \hat{\beta}_1 \) are derived from

\[(6.13) \quad X'_1X_1\hat{\beta}_1 = X'_1y.\]

Clearly,

\[(6.14) \quad \hat{f}^{(1)}_{i} = y_{i}.\]

and

\[(6.15) \quad \hat{f}^{(2)}_{ij} = y_{ij} - y_{i}.\]
where the (·) denotes taking the mean over that subscript.BLUE's of estimable functions of $f_{i}^{(2)}$ alone are made up of functions of $f_{i}^{(2)}$. Since there are no functions of $f_{i}^{(1)}$ alone that are estimable, it is best to calculate BLUE's of confounded functions of the form

$$\lambda_{i}^{1} \beta_{1} + \lambda_{i}^{2} \beta_{2},$$

where $\lambda_{i}^{2} \beta_{2}$ is conditionally zero (conditioned on expression 6.10).

Clearly, any linear function of the observations $a'y$ where $a \in C(X_{1})$ is a conditional BLUE of the parameters $f_{i}^{(1)}$ alone.

The analysis of variance can be displayed as in Table 6.1. The effect of the mean $\mu$ can be removed by subtracting the quantity $y'My$, where $M = \frac{1}{rst} \sum_{rst} y' \sum_{rst}$, from the sum of squares due to factor $F_{1}$.

Table 6.1. Balanced complete nested experiment.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ($\mu$)</td>
<td>1</td>
<td>$y'My$</td>
</tr>
<tr>
<td>Factor $F_{1}$</td>
<td>$s-1$</td>
<td>$y'(M_{1}-M)y$</td>
</tr>
<tr>
<td>Factor $F_{2}$</td>
<td>$s(t-1)$</td>
<td>$y'(M_{2}-M)y$</td>
</tr>
<tr>
<td>Error</td>
<td>$st(r-1)$</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>$rst$</td>
<td>$y'y$</td>
</tr>
</tbody>
</table>
2. **The general case**

For the case $n_{ij} \geq 0$ the model can still be written

\[(6.17)\]
\[y = x_1 \beta_1 + x_2 \beta_2 + \varepsilon,\]

with $C(x_1) \subset C(x_2)$. The two sets of conditions

\[(6.18)\]
\[\sum_{i=1}^{t} \sum_{j=1}^{(2)} i_{ij} = 0 \text{ for all } i\]

and

\[(6.19)\]
\[x_1' x_2 \beta_2 = 0\]

are no longer the same conditions. Although imposing either set to obtain unique solutions to the normal equations will give identical analysis of variance tables, the interpretations given to estimable functions is different.

First, consider conditions 6.19 applied to the reduced normal equations for $\beta_2$ eliminating $\beta_1$

\[(6.20)\]
\[x_2' R_1 x_2 \beta_2 = x_2' R_1 y.\]

Just as in the case when $n_{ij} = r$, a set of solutions $\hat{\beta}_2$ is given by

\[(6.21)\]
\[x_2' x_2 \hat{\beta}_2 = x_2' R_1 y,\]
and the corresponding solutions \( \hat{\beta}_1 \) are given by

\[
(6.22) \quad X'_1X'_1\hat{\beta}_1 = X'_1y .
\]

Since \( C(X_1) \subset C(X_2) \), it follows that there are no estimable functions of \( \beta_1 \) alone; therefore, we resort to conditional unbiasedness again. That is, the function \( a'y \) is a conditional BLUE of \( \lambda'_1\beta_1 \) if \( E[a'y] = \lambda'_1\beta_1 + n'_1X'_1\beta_2 \) for some vector \( n'_1 \). Clearly, for any vector \( a \) such that \( a \in C(X_1) \), the function \( a'y \) is a conditional BLUE of \( \lambda'_1\beta_1 \) for some vector \( \lambda'_1 \). Any function \( b'y \) such that \( b \in C(R_1X_2) \) is BLUE for its expectation, which is a function of \( \beta_2 \) alone.

Now consider conditions 6.18. These conditions have an advantage over conditions 6.19 in that they are independent of the data. The solutions obtained by imposing these conditions when there are no empty cells and when a mean \( \mu \) is included in the model are

\[
(6.23) \quad \hat{\mu} = y_{..} ,
\]

\[
(6.24) \quad \hat{f}^{(1)}_i = y_{i..} - y_{..} ,
\]

\[
(6.25) \quad \hat{f}^{(2)}_{ij} = y_{ij} - y_{i..} ,
\]

where \( y_{ij} = \sum_k y_{ijk}/n_{ij} \), \( y_{i..} = \frac{1}{t} \sum_j y_{ij} \), \( y_{..} = \frac{1}{st} \sum_{ij} y_{ij} \), and \( y_{ij} = \frac{1}{s_i} \sum_j y_{ij} \). The functions

\[
(6.26) \quad \hat{f}^{(1)}_i - \hat{f}^{(1)}_i
\]

are conditional BLUE's of \( f^{(1)}_i - f^{(1)}_i \), where the conditioning is upon 6.18.
When there are missing cells, say the 1,1-th cell, it becomes obscure, as it did in the previous chapter, as to how the condition

\[(6.27) \sum_{i,j} f_{ij}^{(2)} = 0\]

is imposed on the solutions.

B. Tests of Hypotheses

The analysis of variance table for general \((n_{k,j} \geq 0)\) two-factor nested experiment is given in Table 6.2. The effect of the mean has been subtracted from the sum of squares attributable to factor \(F_1\). The matrix \(M\) is the orthogonal projection operator on \(C(1_i)\) where \(n = \sum n_{ij}\) and is equal to \(\frac{1}{n} J_n\).

Table 6.2. General two-factor nested experiment.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ((\mu))</td>
<td>1</td>
<td>(y'M'y)</td>
</tr>
<tr>
<td>Factor (F_1)</td>
<td>(r(X_1) - 1)</td>
<td>(y'(M_1-M)y)</td>
</tr>
<tr>
<td>Factor (F_2)</td>
<td>(r(X_2) - r(X_1))</td>
<td>(y'(M_2-M_1)y)</td>
</tr>
<tr>
<td>Error</td>
<td>(n - r(X_2))</td>
<td>(y'(I-M_2)y)</td>
</tr>
<tr>
<td>Total</td>
<td>(n)</td>
<td>(y'y)</td>
</tr>
</tbody>
</table>

From the fact that \(C(1_n) \subseteq C(X_1) \subseteq C(X_2)\), it follows that the product of any two matrices in the quadratic forms of Table 6.2 is null. That is,
It therefore follows that the quadratic forms in Table 6.2 are independently distributed. Their expectations and expected mean square can be calculated for the random effects, fixed effects, and mixed effects models.

For the random effects models $\beta_1$ and $\beta_2$ are assumed to be independent random vectors with variances $\sigma^2_{1\text{s}}$ and $\sigma^2_{2\text{st}}$. This gives $E[y'y']$ as

$$E[y'y'] = \sigma^2 I + \sigma^2_1 X_1' X_1 + \sigma^2_2 X_2' X_2'.$$

Using the results of Chapter IV, the expected value of the sum of squares for $F_1$ is

$$E[SS(F_1)] = \sigma^2 (r(X_1) - 1) + \sigma^2_1 tr(X_1'(M_1 - M)X_1) + \sigma^2_2 tr(X_2'(M_1 - M)X_2).$$
The expected value of the sum of squares for $F_2$ is

\[(6.31) \quad \mathbb{E}[SS(F_2)] = \sigma^2 (r(X_2) - r(X_1)) + \sigma_2^2 \text{tr}(X_2'(I - M_1)X_2) .\]

The expected mean squares corresponding to these sums of squares are displayed in the table below.

Table 6.3. Expected mean squares for the random effects nested model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom*</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor $F_1$</td>
<td>$q_1$</td>
<td>$\sigma^2 + \sigma_1^2 \text{tr}(X_1'(I-M)X_1)q_1 + \sigma_2^2 \text{tr}(X_2'(M_1-M)X_2)/q_1$</td>
</tr>
<tr>
<td>Factor $F_2$</td>
<td>$q_2$</td>
<td>$\sigma^2 + \sigma_2^2 \text{tr}(X_2'(I-M_1)X_2)/q_2$</td>
</tr>
</tbody>
</table>

* $q_1 = r(X_1) - 1$, and $q_2 = r(X_2) - r(X_1)$.

For the fixed effects model we assume $\beta_1$ and $\beta_2$ are fixed vectors of constants, $a$ and $b$. The expected value of $yy'$ is

\[(6.32) \quad \mathbb{E}[yy'] = \sigma^2 I + (X_1a + X_2b)(X_1a + X_2b)' .\]

Therefore, the expected value of the sums of squares for $\beta_1$ and for $\beta_2$ alone are respectively

\[(6.33) \quad \mathbb{E}[SS(F_1)] = \sigma_1^2 + ((X_1a + X_2b)'(M_1-M)(X_1a + X_2b)) ,\]
with \( q_1 \) and \( q_2 \) as defined in Table 6.3.

The expected mean squares corresponding to expressions 6.33 and 6.34 are displayed in the table below.

Table 6.4. Expected mean squares for the fixed effects nested model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor ( F_1 )</td>
<td>( q_1 )</td>
<td>( \sigma^2 + (X_1 a + X_2 b)'(M_1-M)(X_1 a + X_2 b)/q_1 )</td>
</tr>
<tr>
<td>Factor ( F_2 )</td>
<td>( q_2 )</td>
<td>( \sigma^2 + b'X_2'X_2 b/q_2 )</td>
</tr>
</tbody>
</table>

The mixed model yields two distinct cases for a nested factor because the results are different for \( \beta_1 \) fixed and \( \beta_2 \) random from what they are when \( \beta_1 \) is assumed to be random and \( \beta_2 \) fixed.

**Case I.** Assume \( \beta_1 \) is a vector of fixed constants \( a \) and \( \beta_2 \) is a vector of independently distributed random variables \( b \) with mean zero and variance \( \sigma^2_2 \). The appropriate expected mean squares can easily be calculated and are displayed in Table 6.5.
Table 6.5. Expected mean squares for mixed effects model (Case I).

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor $F_1$</td>
<td>$q_1$</td>
<td>$\sigma^2 + \sigma^2_{a} \frac{\text{tr} (X_1(I-M)X_1)}{q_1} + b'X_2'(M_1-M)X_2b/q_1$</td>
</tr>
<tr>
<td>Factor $F_2$</td>
<td>$q_2$</td>
<td>$\sigma^2 + b'X_2'(M_1-M)X_2b/q_2$</td>
</tr>
</tbody>
</table>

In this situation it may also be reasonable to assume that the parameters $\beta_{ij}^{(2)}$ and $\beta_{ij}^{(2)}$ are correlated. Then the vector $\beta_2$ has variance-covariance matrix of the form

\[
V = \begin{pmatrix}
V_1 & \phi \\
\phi & V_2 \\
\end{pmatrix}
\]

Partitioning $X_2$ into $X_2_1, \ldots, X_2_s$ gives the following expression for $E[yy']$.

\[
E[yy'] = \sigma^2 I + (X_1a)(X_1a)' + \sum_{i=1}^{s} X_{2i}V_{i}X_{2i}'.
\]

The expectation of the quadratic form $y' Ay$ can be written as
Furthermore, if it is assumed $V_i$ has the form

\begin{equation}
V_i = v_i I_t + \rho_i J_t
\end{equation}

then expression 6.37 can be written

\begin{equation}
E[y'y] = \sigma^2 \text{tr}(A) + (X_1 a)'A(X_1 a)
\end{equation}

\[ + \sum_{i=1}^{s} \text{tr}(AX_{2i},V_{12_i}) \]

The expected mean squares using expression 6.39 are presented in the following table.
Table 6.6. Expected mean squares for the mixed effects model when the parameters $\beta_2$ have a correlation structure.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor $F_1$</td>
<td>$q_1$</td>
<td>$\sigma^2 + a'X_1'(I-M)X_1a/q_1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+ \sum_{i=1}^{s} v_i \text{tr}(X_{2i}'(I-M)X_{2i})/q_1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+ \sum_{i=1}^{s} \rho_i l'X_i'(1-M)X_i l'/q_1$</td>
</tr>
<tr>
<td>Factor $F_2$</td>
<td>$q_2$</td>
<td>$\sigma^2 + \sum_{i=1}^{s} v_i \text{tr}(X_{2i}R_iX_{2i}')/q_2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+ \sum_{i=1}^{s} \rho_i l'X_i'R_iX_i l'/q_2$</td>
</tr>
</tbody>
</table>

Case II. Assume $\beta_1$ is a vector of random variables with mean $\phi$ and variance $\sigma^2 I_s$, and $\beta_2$ is a vector of known constants $b$. Then the expectation of $yy'$ is

\begin{equation}
E[yy'] = \sigma^2 I + \sigma^2 X_1X_1' + (X_3b)(X_3b)' .
\end{equation}

The expected mean squares for the quantities in Table 6.2 are given in Table 6.7.
Table 6.7. Expected mean squares for the mixed effects model (Case II).

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Expected Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor $F_1$</td>
<td>$q_1$</td>
<td>$\sigma^2 + \sigma_1^2 \text{tr}(X_1'(I-M)X_1)/q_1 + b'M_1X_2b/q_1$</td>
</tr>
<tr>
<td>Factor $F_2$</td>
<td>$q_2$</td>
<td>$\sigma^2 + b'X_2'X_2b/q_2$</td>
</tr>
</tbody>
</table>
VII. MULTIFACTOR EXPERIMENTS

In this chapter the results of Chapter IV are extended to models with more than two partitions. Natural partitions of estimable functions and their corresponding BLUE's are used to partition the estimation space \( C[X] \).

A. Estimation

The partitioned linear model with \( p \) partitions will be written as

\[
y = \sum_{i=1}^{p} X_i \beta_i + \epsilon,
\]

where \( X_i \) and \( \beta_i \) have dimensions \( n \times p_i \) and \( p_i \times 1 \), respectively, and \( \epsilon \) is a vector of random variables with mean zero and variance \( \sigma^2 I \). In the development that follows extensive use will be made of the following operators. Define:

\[
P[i_1, i_2, \ldots, i_s] \quad \text{to be the orthogonal projection operator on the column space of the matrix} \quad (X_{i_1}, X_{i_2}, \ldots, X_{i_s}),
\]

\[
M[i_1, i_2, \ldots, i_s] \quad \text{to be the orthogonal projection operator on intersection of the column spaces of} \quad X_{i_1}, X_{i_2}, \ldots, \text{and} X_{i_s},
\]

\[
R[i_1, i_2, \ldots, i_s] \quad \text{to be the orthogonal projection operator on the complement of the column space of} \quad (X_{i_1}, X_{i_2}, \ldots, X_{i_s}).
\]
Let $S$ denote a particular set of subscripts

(7.6) \[ S = \{i_1, i_2, \ldots, i_s\} \]

then the following hold,

1) \[ R[S] = I - P[S] \]

ii) \[ N[S] = I - M[S] \]

The order in which the subscripts appear as well as the number of times a subscript appears within the brackets is immaterial. The following additional properties will be stated without proof.
iii) $P[i] = M[i,i]$

iv) $R[i] = N[i,i]$


If $S_1$ and $S_2$ are sets of subscripts such that $S_1 \subset S_2$ then


x) $P[S_1]P[S_2] = P[S_2]P[S_1] = P[S_1]$

xi) $R[S_1]R[S_2] = R[S_2]R[S_1] = R[S_2]$

xii) If $R[S_1]z = \phi$ for some vector $z$, then $R[S_2]z = \phi$.

xiii) If $N[S_1]z = \phi$ for some vector $z$, then $N[S_2]z = \phi$.

xiv) If $P[S_1]z = z$ for some vector $z$, then $P[S_2]z = z$.

xv) If $M[S_2]z = z$ for some vector $z$, then $M[S_1]z = z$.

Let $S$ be the complete set of indices

(7.7) $S = \{1,2,\ldots,p\}$

then define for any set of indices $S_1$ a set $\bar{S}_1$ which is the complement of $S_1$, i.e. $\bar{S}_1$ is the set of indices in $S$ that are not in $S_1$. Also let $\bar{i}$ denote all subscripts except $i$.

Using the operators and notation defined above, a generalization of Theorem 4.1 can be established. Consider the following groups of column spaces:
Group 1 : $C(\mathbb{R}[\mathbb{I}]X_i)$, $i = 1, 2, \ldots, p,$

Group 2 : $C(\mathbb{R}[\mathbb{J}]X_i) \cap C(\mathbb{R}[\mathbb{J}]X_j)$, $i, j = 1, 2, \ldots, p,$

Group 3 : $C(\mathbb{R}[\mathbb{J}X_i]) \cap C(\mathbb{R}[\mathbb{J}X_j]) \cap C(\mathbb{R}[\mathbb{J}X_k])$, $i, j, k = 1, 2, \ldots, p,$

(7.8) 

Group $p-1$: $C(\mathbb{R}[i]X_i) \cap \ldots \cap C(\mathbb{R}[i]X_{i-1}) \cap \ldots \cap C(\mathbb{R}[i]X_p)$, $i = 1, 2, \ldots, p,$

Group $p$: $C(\mathbb{X}_1) \cap C(\mathbb{X}_2) \cap \ldots \cap C(\mathbb{X}_p)$.

Since $P[S]$ has the form

(7.9) $P[S] = (X_{i_1} \ldots X_{i_s})(X_{i_1} \ldots X_{i_s})'(X_{i_1} \ldots X_{i_s})' (X_{i_1} \ldots X_{i_s})'$

and $M[S]$ has the form

(7.10) $M[S] = X_{i_1}A(A'X_{i_1}X_{i_1}A)^{-1}A'X_{i_1}$

for some matrix $A$, it is clear that a vector $z$ in any of the above column spaces can be written as

(7.11) $z = \sum_{i=1}^{P} X_i \rho_i$,

for some vectors $\rho_1, \ldots, \rho_p$. Therefore a vector in any of these column spaces is BLUE for its expectation. Let $a$ be a vector in one of the first group of column spaces, i.e.
Consider

\[(7.12) \quad a \in C(R[k]X_k) .\]

Expression \(7.12\) now becomes

\[(7.13) \quad E[a'y] = a'\ E(y) = \sum_{i=1}^{p} a'X_i\beta_i .\]

Therefore \(a'y\) is BLUE for an estimable function of \(\beta_k\) alone.

Next let \(a\) be a vector contained in one of the column spaces in the second group, i.e.

\[(7.14) \quad a = R[k]X_k\rho .\]

\[(7.15) \quad E[a'y] = \sum_{i=1}^{p} \rho'X_kR[k]X_k\beta_i = \rho'X_kR[k]X_k\beta_k = a'X_k\beta_k = \lambda'\beta_k .\]

Since \(a \in C(R[k]X_k)\), there exists a vector \(\rho\) such that

\[(7.16) \quad a \in C(R[ij]X_i) \cap C(R[ij]X_j) .\]
The expected value of $a'y$ is

\begin{equation}
E[a'y] = \sum_{k=1}^{p} a'X_k \beta_k
\end{equation}

\begin{align*}
&= \eta_1^t X_1 R[\bar{1}] X_1 \beta_1 + \eta_2^t X_1 R[\bar{1}] X_j \beta_j \\
&= \lambda_1 \beta_i + \lambda_2 \beta_j .
\end{align*}

Therefore $a'y$ is BLUE for an estimable function of $\beta_i$ and $\beta_j$. The following theorem will be used to show that $\lambda_1 \beta_i$ and $\lambda_2 \beta_j$ are not individually estimable.

**Theorem 7.1** The column spaces listed in expression 7.8 are disjoint.

**Proof** First consider those spaces represented by

$$C(R[\bar{i}]X_i) \quad i = 1, 2, \ldots, \rho .$$

Suppose the vector $a$ is in both $C(R[\bar{i}]X_i)$ and $C(R[\bar{j}]X_j)$, then there exists vectors $\eta_1$ and $\eta_2$ such that

$$a = R[\bar{i}]X_i \eta_1 = R[\bar{j}]X_j \eta_2 .$$

The quantity $a'a$ can be written as

$$a'a = \eta_1^t X_1 R[\bar{1}] R[\bar{1}] X_1 \eta_1 .$$

Since $R[\bar{1}]$ is symmetric and idempotent,
Therefore, within the first group the column spaces are disjoint.

Now, consider a vector \( a \) that is contained in two distinct column spaces of the second group. Applying the same type of argument as above gives

\[
a \in C(\mathbf{R}[\overline{ij}]X_1) \cap C(\mathbf{R}[\overline{ij}]X_j)
\]

and

\[
a \in C(\mathbf{R}[\overline{nm}]X_n) \cap C(\mathbf{R}[\overline{nm}]X_m),
\]

where at least one of the subscripts \( i \) or \( j \) is different from the subscripts \( n \) and \( m \). Assume \( i \neq n \) and \( i \neq m \). Then

\[
a = \mathbf{R}[\overline{ij}]X_i \eta_1
\]

and

\[
a = \mathbf{R}[\overline{nm}]X_n \eta_2
\]

for some vectors \( \eta_1 \) and \( \eta_2 \). The quantity \( a'a \) can be written as

\[
a'a = \eta_1^T X^i \mathbf{R}[\overline{ij}]' \mathbf{R}[\overline{ij}]X_i \eta_1
\]

\[
= \eta_1^T X^i \mathbf{R}[\overline{ij}]X_i \eta_1
\]

\[
= \eta_1^T X^i \mathbf{R}[\overline{nm}]X_n \eta_2
\]

\[
= \phi.
\]
Using the fact that $R[S]X_i = \phi$ for all $X_i$ such that $i \notin S$, the procedure above can be repeated for all the groups of column spaces in expression (7.8). This procedure can also be used to show the disjointedness of column spaces in different groups.

Consider any vector $a$ from one of the column spaces in group $i$ that is also a vector in one of the column spaces of group $j$ with $i \neq j$.

Then $a$ can be expressed as

$$a = R[S_i]X_k \eta_1$$

$$= R[S_j]X_k \eta_2$$

for some vectors $\eta_1$ and $\eta_2$ where $k$ is any subscript in $S_i$ and $l$ is any vector in $S_j$. Thus

$$a'\ a = \eta_1'X_i R[S_i]'R[S_i]X_k \eta_1$$

$$= \eta_1'X_i R[S_i]X_k \eta_1$$

$$= \eta_1'X_i R[S_j]X_k \eta_2$$

Since the sets of subscripts $S_i$ and $S_j$ are not the same set, there is a subscript $h$ such that $h \in S_i$ and $h \in S_j$. Letting $k = h$ gives
\[ a' a = \eta_1^j x_R[S_j] x_2 n_2 \]
\[ = \eta_1^j(\phi) x_2 n_2 \]
\[ = \phi . \]

This completes the proof of Theorem 7.1.

We can associate with any set of subscripts
\[ S = \{i_1, i_2, \ldots, i_s\} , \]
the space
\[ C(S) = \bigcap_{i \in S} C[R[S_j]x_1] \]
with \( R[\phi] = I \). It can easily be seen that there are \( 2^P - 1 \) distinct sets \( S \). Therefore, \( C(X_1, \ldots, X_p) \) can be partitioned into at least \( 2^P - 1 \) disjoint subspaces. The following lemma will be useful for establishing that the \( 2^P - 1 \) subspaces \( C(S) \) in fact span \( C(X_1, \ldots, X_p) \).

**Lemma 7.1** Let \( B \) be any basis for \( \bigcap_{i \in S} C[X_1] \). Then the two spaces
\[ \bigcap_{i \in S} C[R[S_j]x_1] \]
and
\[ C[R[S]B] \]
are the same.
Proof. Let \( a \) be any vector contained in \( C(R[S]B) \), then \( a \) can be written as

\[ a = R[S]X_i \eta_i \]

for all \( i \in S \) and for some set of vectors \( \eta_i \). This implies that

\[ a \in \bigcap_{i \in S} (C(R[S]X_i)) \]

Therefore,

\[ C(R[S]B) \subseteq \bigcap_{i \in S} C(R[S]X_i) \]

Let \( b \) be any vector in \( \bigcap_{i \in S} C(R[S]X_i) \). This implies that for any \( i \in S \) there exists a vector \( \eta_i \) such that

\[ b = R[S]X_i \eta_i \]

Without loss of generality we can assume that \( S \) contains \( k \) elements, \( 1, 2, \ldots, k \). For \( i = 1 \) and \( 2 \),

\[ b = R[S]X_1 \eta_1 \]

\[ = R[S]X_2 \eta_2 \]

Therefore,
\[ R[\bar{S}]x_1\eta_1 - R[\bar{S}]x_2\eta_2 = \phi , \]
\[ R[\bar{S}](x_1\eta_1 - x_2\eta_2) = \phi , \]
\[ x_1\eta_1 - x_2\eta_2 = p[\bar{S}](x_1\eta_1 - x_2\eta_2) , \]
and finally

\[ x_1\eta_1 = x_2\eta_2 + p[\bar{S}](x_1\eta_1 - x_2\eta_2) . \]

Since the last term is a vector in \( C(x_{k+1}, \ldots, x_p) \), we can write

\[ x_1\eta_1 = x_2\eta_2 + \rho , \]

where \( \rho \in C(x_{k+1}, \ldots, x_p) \). Equivalently,

\[ b = R[\bar{S}]\eta , \]

where \( \eta \in [C(x_1) \cap C(x_2, x_{k+1}, \ldots, x_p)] \). Any such vector \( \eta \) satisfies

\[ \eta = \xi_1 + \xi_2 \]

where \( \xi_1 \in [C(x_1) \cap C(x_2)] \) and \( \xi_2 \in [C(x_{k+1}, \ldots, x_p)] \). Because for any \( \xi_2 , R[\bar{S}] = \phi , \) therefore

\[ b = R[\bar{S}]\xi_1 \]
where $\xi_1 \in [C(X_1) \cap C(X_2)]$. Now treating $C(X_1) \cap C(X_2)$ as a space $Z$ we show in the same way that

$$b = R[S]\xi_3$$

where

$$\xi_3 \in [C(Z) \cap C(X_3)]$$

or

$$\xi_3 \in [C(X_1) \cap C(X_2) \cap C(X_3)]$$

Hence

$$b = R[S]n,$$

where $n \in [C(X_1) \cap C(X_2) \cap C(X_3)]$. This procedure can be continued for all $i \in S$ to give the result

$$\bigcap_{i \in S} C[R[S]X_i] \subset C[R[S]B],$$

and therefore the two spaces above are identical.

Next we will determine the dimension of $C(S)$ by establishing the rank of the matrix $R[S]B$. Since for any other basis $B_0$ for $\bigcap_{i \in S} C(X_i)$ we can write

$$B = B_0D$$
for some matrix $D$, and any vector $\rho$ in $C(R[\tilde{S}]B_0)$ can be written as

$$\rho = R[\tilde{S}]B_0\eta$$

for some $\eta$, it follows that there exists a vector $\gamma$ such that

$$\rho = R[\tilde{S}]B_0\gamma$$

Therefore, construct $B_0$ as follows. Let $B_0 = (B_1, B_2)$ where $B_1$ is a full-rank basis for $C(\{\tilde{S}\} \cap C(B)$ and $B_2$ is its extension to a full-rank basis for $C(B)$. Then $R[\tilde{S}]B_0$ forms a basis for $C(R[\tilde{S}]B)$. We can write

$$R[\tilde{S}]B_0 = (\phi, R[\tilde{S}]B_2) .$$

Assume that the columns of $R[\tilde{S}]B_2$ are not linearly independent. Then there exists a nonzero vector $a$ such that

$$R[\tilde{S}]B_2a = \phi .$$

This implies that

$$B_2a = P[\tilde{S}]B_2a .$$

Because of the construction of $B_2$, the vector $B_2a$ must be nonzero. However this implies that $B_2a$ is a vector in $C(\{\tilde{S}\})$ which is impossible. Therefore, no such vector $a$ exists, and the columns of $R[\tilde{S}]B_2$ are linearly independent. This gives
\[ (7.18) \quad \text{rank}(R[S]B) = \text{rank}(R[S]B_0) \]
\[ = \text{rank}(B) - \dim[C[P[S]] \cap C(B)] \]
\[ = \dim[C(B)] - \dim[C(P[S]) \cap C(B)]. \]

From Lemma 7.1 and expression 7.18, the following lemma can be proved.

**Lemma 7.2**  Let \( J \) denote the collection of \( 2^p - 1 \) possible subsets \( S \) of the first \( p \) integers that yield distinct spaces \( C(S) \) for \( S \in J \). Then the following holds

\[ (7.19) \quad \dim[C(X_1, ..., X_p)] = \sum_{S \in J} \dim[C(S)]. \]

**Proof**  For convenience we will construct the class \( J \) so that it contains exactly \( 2^p - 1 \) sets \( S \), where each set has the form

\[ (7.20) \quad S = \{i_1, ..., i_k\}, \]

with \( i_1 < i_2 < ... < i_k \). The dimension of \( C(X_1, ..., X_p) \) can be written as

\[ (7.21) \quad \dim[C(X_1, ..., X_p)] = \{\dim[C(X_1)] - \dim[C(X_1) \cap C(X_2, ..., X_p)]\} \]
\[ + \{\dim[C(X_2)] - \dim[C(X_2) \cap C(X_3, ..., X_p)]\} \]
\[ + \{\dim[C(X_p)] - \dim[C(X_p) \cap C(X_1)]\} \]
\[ + \{\dim[C(X_p)]\}. \]
From expression 7.18 it follows that
\[ \dim[C(X_1)] - \dim[C(X_1) \cap C(x_2, \ldots, x_p)] = \dim[C(1)]. \]

Now consider the second line of expression 7.21, with quantity, \( \dim[C(X_2) \cap C(x_1, x_3, \ldots, x_p)] \), added and subtracted,

\[ (7.22) \quad \dim[C(X_2)] - \dim[C(X_2) \cap C(X_1, x_3, \ldots, x_p)] + \dim[C(X_2) \cap C(X_1, x_3, \ldots, x_p)] - \dim[C(X_2) \cap C(x_3, \ldots, x_p)]. \]

From expression 7.18 it follows that the first line of 7.22 is \( \dim[C(2)] \).

Using the fact that for arbitrary matrices \( A \) and \( B \), each with \( n \) rows,

\[ (7.23) \quad \dim[C(A,B)] = \dim[C(A)] + \dim[C(B)] - \dim[C(A) \cap C(B)]. \]

Letting \( C(A) = C(X_2) \cap C(X_1) \) and \( C(B) = C(X_2) \cap C(x_3, \ldots, x_p) \), it follows that

\[ (7.24) \quad \dim[C(X_2) \cap C(X_1, x_3, \ldots, x_p)] = \dim[C(X_2) \cap C(x_1)] + \dim[C(X_2) \cap C(x_3, \ldots, x_p)] - \dim[C(X_2) \cap C(x_1) \cap C(x_3, \ldots, x_p)]. \]
Substituting expression 7.24 in 7.22 gives

$$\dim[C(2)] + \dim[C(x_1) \cap C(x_2)]$$

$$- \dim[C(x_1) \cap C(x_2) \cap C(x_3, \ldots, x_p)]$$

$$+ \dim[C(x_2) \cap C(x_3, \ldots, x_p)]$$

$$- \dim[C(x_2) \cap C(x_3, \ldots, x_p)]$$

$$= \dim[C(2)] + \dim[C(1,2)].$$

A similar type of argument can be applied to the third line of 7.21 to show that

$$\dim[C(x_3)] - \dim[C(x_3) \cap C(x_4, \ldots, x_p)]$$

$$= \dim[C(3)] + \dim[C(1,3)] + \dim[C(2,3)]$$

$$+ \dim[C(1,2,3)].$$

In general, the k-th line of expression 7.21 gives

$$\sum \dim[C(S)]$$

such that the maximum element of S is equal to k. Therefore, every set $S \in J$ is represented in one and only one of the lines in expression 7.21, and

$$\sum \dim[C(x_1, \ldots, x_p)] = \sum \dim[C(S)].$$

(7.25)
Lemmas 7.1 and 7.2 therefore establish the following extension of Theorem 4.2.

**Theorem 7.2** Given a partitioned linear model with \( p \) partitions, any linear function of the observations \( a'y \) that is BLUE for its expectation can be written

\[
a'y = \sum_{i=1}^{C_p} \rho_{ij}y + \ldots + \sum_{i=1}^{C_p} \rho_{ip}y,
\]

where \( C_i = \binom{p}{i}, \) \( i = 1, 2, \ldots, p \) and \( \rho_{ij} \) is a vector contained in the \( i \)-th column space of the \( j \)-th group. In other words

\[
\rho_{ij}y
\]

is BLUE for an estimable function in \( \beta_i \) alone,

\[
\rho_{i2}y
\]

is BLUE for an estimable function in two of the sets of parameters, say \( \beta_k \) and \( \beta_{k'} \), and so on until

\[
\rho_{ip}y
\]

is BLUE for an estimable function in all of the sets of parameters. The following extension of the definitions of confounding given in Chapter IV will be of use.

**Definition** The \( s \) linear functions \( \lambda_1\beta_{i1}, \lambda_2\beta_{i2}, \ldots, \lambda_s\beta_{is} \) are said to be *completely confounded* whenever the function
is estimable and has BLUE of the form \( \delta'y \) where \( \delta \) is a vector such that
\[
\delta \in C(\bigcap_{i \in S} R[\vec{S}]x_i)
\]
and \( S = \{i_1, i_2, \ldots, i_s\} \). The function \( \gamma \) can also be said to be \textit{inseparable}.

\textbf{Definition} The linear function \( \lambda_{i_1} \beta_{i_1} + \ldots + \lambda_{i_s} \beta_{i_s} \) is said to be a separable estimable function if it is estimable and its BLUE can be written as \( \delta'y \) where \( \delta \) can be written nontrivially as the sum
\[
\delta = \delta_1 + \delta_2
\]
with \( \delta_1 \in C(S), \delta_2 \notin C(S) \), where \( S = \{i_1, \ldots, i_s\} \).

\textbf{B. The Maximum Number of Partitions}

One of the most apparent drawbacks of this partitioning process seems to be the inordinately large number of partitions generated. In this section it is demonstrated that in most cases many of these spaces are null. For example, suppose there exists subscripts \( i \) and \( j \) (\( i \neq j \)) such that
\[
C(x_i) \subset C(x_j).
\]
Then it is clear that \( R[\vec{S}]x_i = \emptyset \) for any set \( S \) such that \( j \notin S \). Therefore, for all sets of indices \( S \) such that \( j \notin S \) and \( i \in S \),
There are $2^{P-1} - 1$ sets $S$ that do not contain $j$, and of these sets, $2^{P-2}$ of them contain $i$. Therefore, whenever one of the factors in a linear model nests another factor, there can be no more than

$$2^P - 2^{P-2} - 1$$

spaces $C(S)$. 

Suppose there exists subscripts $i, j, k$ such that

$$C(X_k) \subseteq C(X_i) \subseteq C(X_j),$$

then $C(S) = \emptyset$ for the following cases:

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>j</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>0</td>
<td>X</td>
</tr>
</tbody>
</table>

where $X$ denotes inclusion in a set and $0$ denotes exclusion. There are $2^{P-3}$ sets having each of the four configurations; therefore, there are no more than
spaces $C(S)$ which are non-null. When there are four successively nested factors

$$C(x_{\ell}) \subseteq C(x_k) \subseteq C(x_i) \subseteq C(x_j),$$

the following situations produce null spaces:

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>$k$</th>
<th>$i$</th>
<th>$j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$x$</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>$x$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>$x$</td>
<td>$x$</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>$x$</td>
<td>0</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x$</td>
<td>0</td>
<td>0</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>0</td>
<td>$x$</td>
<td>0</td>
</tr>
<tr>
<td>$x$</td>
<td>0</td>
<td>$x$</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>0</td>
<td>$x$</td>
</tr>
<tr>
<td>$x$</td>
<td>$x$</td>
<td>$x$</td>
<td>0</td>
</tr>
</tbody>
</table>

There are $2^{p-4}$ sets satisfying each of the eleven configurations, so there can be no more than

$$(7.28) \quad 2^p - (11)2^{p-4} - 1$$
spaces \( C(S) \) that are non-null. In general, if \( r \) spaces are successively nested, there will be no more than

\[
2^p - (2^r - r - 1)2^{p-r} - 1
\]

spaces \( C(S) \) that are not empty. When all \( p \) column spaces \( C(X_i) \) \( i = 1, 2, \ldots, p \) are nested, there can be no more than \( p \) spaces \( C(S) \) that are not empty.

Now consider the case when there exists subscripts \( i, j \) and \( k \) such that

\[
C(X_i) \subseteq C(X_k),
\]

and

\[
C(X_j) \subseteq C(X_k).
\]

Then \( C(S) \) will be null for all \( S \) such that \( k \not\in S \) and \( i \in S \) or \( j \in S \). There are \( 2^{p-3} \) sets that satisfy each of the configurations below:

\[
\begin{array}{ccc}
1 & 1 & k \\
0 & X & 0 \\
X & 0 & 0 \\
X & X & 0
\end{array}
\]

Therefore, there are no more than
spaces $C(S)$ that are non-null. Suppose there are $r$ subscripts \{${i_1}...{i_r}$\} such that

$$C(x_i^j) \subseteq C(x_{k})$$

Then $C(S) = \phi$ whenever $k \notin S$ and $i_j \in S$ for at least one $j$. There are $2^r - 1$ non-empty subsets of \{${i_1}...{i_r}$\}. For each of these configurations there are $2^{p-r-1}$ sets $S$ that contain exactly one of these subsets and have $k$ absent. Therefore, in general, there can be no more than

$$(7.31) \quad 2^p - (2^r - 1)2^{p-r-1} - 1$$

spaces $C(S)$ that are non-empty. The case $r = 2$ will arise when the matrix $X_k$ represents the interaction of $X_i$ and $X_j$. Of course, in any case the number of non-null subspaces can never exceed the rank $(X_1,...,X_p)$.

Suppose that

$$C(x_i^j) \subseteq C(x_{k})$$

for all $i = 1,2,...,p$. Then $C(S) = \phi$ for all $S$ such that $i \in S$ except for the set $S = \{1,2,...,p\}$. There are $2^{p-1}$ sets $S$ that contain $i$, one of them being the set with all subscripts present. Therefore, there is a maximum of
\[ 2^p - (2^{p-1} - 1) - 1 = 2^p - 2^{p-1} = 2^{p-1} \]

non-null spaces \( C(S) \). This result is particularly applicable since \( X_1 \) usually corresponds to a column vector of ones and forms a one-dimensional subspace of every other partition regardless of unequal numbers or missing cells as long as each of the other matrices \( X_i \) corresponds to one of the terms in the model.

The maximum numbers of non-empty subspaces \( C(S) \) are given in Table 7.1 for some common designs.

**Table 7.1. Maximum number of non-null subspaces \( C(S) \).**

<table>
<thead>
<tr>
<th>Model</th>
<th>Maximum Number of Partitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_{ijk} = \mu_{ij} + \varepsilon_{ijk} )</td>
<td>1</td>
</tr>
<tr>
<td>( y_{ij} = \mu + a_i + \varepsilon_{ij} )</td>
<td>2</td>
</tr>
<tr>
<td>( y_{ijk} = \mu + a_{ij} + \varepsilon_{ijk} )</td>
<td>2</td>
</tr>
<tr>
<td>( y_{ijk} = \mu + a_i + b_j + \varepsilon_{ijk} )</td>
<td>4</td>
</tr>
<tr>
<td>( y_{ijk} = \mu + a_i + b_j + ab_{ij} + \varepsilon_{ijk} )</td>
<td>5</td>
</tr>
<tr>
<td>( y_{ijk} = \mu + a_i + b_{ij} + \varepsilon_{ijk} )</td>
<td>3</td>
</tr>
<tr>
<td>( y_{ijk} = \mu + a_i + b_j + c_{jk} + \varepsilon_{ijk} )</td>
<td>6</td>
</tr>
<tr>
<td>( y_{ijkl} = \mu + a_i + b_j + c_k + \varepsilon_{ijkl} )</td>
<td>9</td>
</tr>
<tr>
<td>( y_{ijkl} = \mu + a_i + b_j + ab_{ij} + c_k + ac_{ik} + bc_{jk} + abc_{ijk} + \varepsilon_{ijkl} )</td>
<td>14</td>
</tr>
</tbody>
</table>
C. Orthogonal Data Situations

The linear model

\[ y = \sum_{i=1}^{p} X_i \beta_i + \epsilon \]  

is said to be of maximal rank if \( \text{rank}(X_1, \ldots, X_p) \) cannot be increased by adding more observations consistent with the model restrictions.

There are cases when the model is of maximal rank and the subclass frequencies are unequal, yet the data still admit an orthogonal decomposition of the estimation space, e.g. proportional subclass frequencies. Such models will be called orthogonal models of maximal rank. A main-effects model that is orthogonal of maximal rank would have the analysis of variance displayed in Table 7.2.

Table 7.2. Main-effects, orthogonal maximal rank model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (( \beta_1 ))</td>
<td>1</td>
<td>( y'(M[1])y )</td>
</tr>
<tr>
<td>Factor ( F_1 (\beta_2) )</td>
<td>( p_2 - 1 )</td>
<td>( y'(M[2] - M[1])y )</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Factor ( F_{p-1} (\beta_p) )</td>
<td>( p_p - 1 )</td>
<td>( y'(M[p] - M[1])y )</td>
</tr>
<tr>
<td>Error</td>
<td>( n - (\sum p_i - p + 1) )</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>( n )</td>
<td>( y'y )</td>
</tr>
</tbody>
</table>
The matrices of the quadratic forms in Table 7.2 are orthogonal, and therefore, each sum of squares and corresponding mean square is independently distributed for $y$ with variance $\sigma^2 I$. The expected mean squares can be calculated by evaluating $E[yy']$ for the hypothesis in question, then

$$E[y'Ay] = \text{tr}(A E[yy']) .$$

Orthogonal models of submaximal rank can also be defined. Model 7.1 is said to be of submaximal rank when rank $(X_1, \ldots, X_p)$ can be increased by adding more observations consistent with the model restrictions. In such a case it is necessary to examine the individual space $C(S)$ generated by the analysis. Therefore, let $J$ be the class of sets $S$ such that if $S \in J$ then $C(S)$ is non-null. For orthogonal submaximal rank models

$$C(S) \perp C(T)$$

for $S \in J$ and $T \in J$. In order to display the analysis of variance and tests of hypothesis more clearly, the elements of $J$ can be enumerated in the following manner. Any set $S \in J$ consists of integers whose values are between 1 and $p$ inclusive. Assign two subscripts $i$ and $j$ to sets $S$ where $i$ denotes the cardinality of the set and $j$ is a number between 1 and the number of sets $S_{ij}$ with cardinality $i$. In other words the elements of $J$ can be listed as
where $k_1$ is the number of sets in $J$ with cardinality $i$. To simplify notation and for use later let $M[S_{ij}]$ be the orthogonal projection operator on $C(S_{ij})$. For the case under discussion $M[S_{ij}] = \bigoplus_{k \in S_{ij}} M[k]$. The analysis of variance table can be presented in sections as in Table 7.3.
Table 7.3. Analysis of variance for an orthogonal submaximal rank model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects involving $\beta_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{11}$</td>
<td>$\dim(C(S_{11}))$</td>
<td>$y'M[S_{11}]y$</td>
</tr>
<tr>
<td>$S_{12}$</td>
<td>$\dim(C(S_{12}))$</td>
<td>$y'M[S_{12}]y$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$S_{1k'_1}$</td>
<td>$\dim(C(S_{1k'_1}))$</td>
<td>$y'M[S_{1k'_1}]y$</td>
</tr>
<tr>
<td>Effects involving $\beta_i$ and $\beta_j$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{21}$</td>
<td>$\dim(C(S_{21}))$</td>
<td>$y'M[S_{21}]y$</td>
</tr>
<tr>
<td>$S_{22}$</td>
<td>$\dim(C(S_{22}))$</td>
<td>$y'M[S_{22}]y$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$S_{2k_2}$</td>
<td>$\dim(C(S_{2k_2}))$</td>
<td>$y'M[S_{2k_2}]y$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>Effects involving all $\beta_i$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{p1}$</td>
<td>$\dim(C(S_{p1}))$</td>
<td>$y'M[S_{p1}]y$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - \dim(C(X))$</td>
<td>difference</td>
</tr>
<tr>
<td>Total</td>
<td>$n$</td>
<td>$y'y$</td>
</tr>
</tbody>
</table>
Since the mean is included in $S_{pl}$, it may be desirable to partition $y'M[S_{pl}]y$ into two parts, one for the mean and one for the remainder. Because $C[S_{ij}]$ and $C[S_{lk}]$ for $i \neq \ell$ or $j \neq k$ we have

$$(7.34) \quad M[S_{ij}] M[S_{lk}] = \phi ,$$

which is sufficient for the quadratic forms $y'M[S_{ij}]y$ and $y'M[S_{lk}]y$ to be independently distributed when $\text{Var}(y) = \sigma^2 I$. Expected mean squares can be derived using 7.33.

The decision whether or not to pool some of the sums of squares in Table 7.3 should be made on the basis of the experimental material.

D. Non-orthogonal Data Situations

In this section the general linear classificatory model is discussed. To aid in developing an understanding of the analysis of variance for this type of model, we will first examine a geometrical interpretation of the analysis.

1. The geometry of the non-orthogonal analysis of variance

Many text books on linear models portray least squares model fitting as in Figure 7.1.

![Figure 7.1. Linear regression](image-url)
The analysis of variance table is nothing more than a display of the distances squared that are displayed in Figure 7.1. Because of the Pythagorean theorem and the fact that two of the distances displayed in Figure 7.1 are orthogonal, the sums of squares add up to the total. The vector \( \hat{y} = X\hat{\beta} \) is the orthogonal projection of the vector \( y \) onto the estimation space \( C(X) \). This space \( C(X) \) characterizes BLUE's of estimable functions of \( \beta \). The vector \( y - X\hat{\beta} \) is the difference between \( y \) and \( \hat{y} \), the distance from \( y \) to \( \hat{y} \), or the orthogonal projection of \( y \) on the complement of \( C(X) \). This complement is commonly called the error space. The analysis of variance generally involves the decomposition of \( C(X) \) into several parts, i.e. \( C(X) = C(S_1, X_2, \ldots, X_p) \). If the parts are orthogonal, a more general form of the Pythagorean theorem applies and the squared distances displayed in the analysis of variance table still sum to the squared length of \( y \). When the subspaces of \( C(X) \) are not orthogonal, the squared distances in the analysis of variance table do not sum to \( y'y \).

The interpretation to be given to the decomposition presented here is as follows. The space of all vectors \( a \) such that \( a'y \) is BLUE of its expectation is partitioned according to \( E[a'y] \). The space spanned by those vectors \( a \) such that

\[
(7.35) \quad E[a'y] = \sum_{i \in S_{ij}} \lambda_i \beta_i,
\]

where any partial sum of 7.36 is not estimable, makes up one partition \( C(S_{ij}) \) of \( C(X) \). The vector
where \( M[S_{ij}] \) denotes the orthogonal projection operator on \( C(S_{ij}) \), is a projection of \( y \) such that the distance from \( y \) to \( C(S_{ij}) \) is a minimum. The quantity \( y'M[S_{ij}]y \) is square of the length of projection of \( y \) onto \( C(S_{ij}) \). The projected vectors \( y_{ij} = M[S_{ij}]y \) for all values of \( i \) and \( j \) are not orthogonal. The cases when these are orthogonal were discussed in previous sections. The next section discusses cases when only some of the \( y_{ij} \) are orthogonal.

2. **Partial orthogonality**

Consider the space \( C(S) \) where \( S = \{1,2,\ldots,p\} \), that is, the intersection of all \( X_i \). This space is orthogonal to all \( C(I) \) since any vector \( a \) in \( C(S) \) can be written as

\[
(7.37) \quad a = X_i \eta
\]

and any vector \( b \) in \( C(I) \) can be written

\[
(7.38) \quad b = \mathbb{R}[i] \rho
\]

therefore, \( a'b = \phi \). This can be extended to the following theorem.

**Theorem 7.3** Let \( S_1 \) and \( S_2 \) be any sets of subscripts such that \( S_1 \subseteq S_2 \) and \( S_1 \neq S_2 \), then the spaces \( C(S_1) \) and \( C(S_2) \) are orthogonal.
Proof Since $S_1 \neq S_2$ there exists at least one subscript $i$ such that $i \in S_2$ but $i \notin S_1$. Any vector $a$ in $C(S_2)$ can be written

$$(7.39) \quad a = R[\bar{S}_2]x_1 \eta$$

for some vector $\eta$. Any vector $b$ in $C(S_1)$ can be written

$$(7.40) \quad b = R[\bar{S}_1] \rho$$

for some vector $\rho$. From property xi

$$(7.41) \quad R[\bar{S}_2] R[\bar{S}_1] = R[\bar{S}_1] .$$

This gives

$$a' b = \eta' x'_1 R[\bar{S}_2] R[\bar{S}_1] \rho$$

$$= \eta' x'_1 R[\bar{S}_1] \rho$$

$$= \phi$$

since $x'_1 R[\bar{S}_1] = \phi$, which establishes the orthogonality property.

Theorem 7.3 establishes a property that will be useful in displaying the analysis of variance and also for proving the following corollary.

Corollary 7.1 The BLUE of the completely confounded function

$$\sum_{i \in S_1} \lambda_i \beta_i$$
is uncorrelated with any estimable function of the form

$$\sum_{j \in S_2} \lambda_j^g \beta_j$$

when $S_1 \subset S_2$.

The above corollary will be used in a subsequent section for calculating the variances of BLUE's of estimable functions.

E. Decomposition of Sums of Squares

In this section the regression sum of squares $\hat{y}'X'H\hat{y}$ or $\hat{y}'\hat{y}$ will be partitioned into a sum of sums of squares where each partition is associated with one of the spaces $C(S)$. Recall that $M(S)$, the orthogonal projection operator on $C(S)$, has the form


Since $X_i'X(S) = X_i'\hat{y} = X_i'y$ for any $X_i, i = 1, \ldots, p$, the following holds,

$$\hat{y}'X'M[S]X(S) = \hat{y}'M[S]\hat{y} = y'M[S]y$$

for all sets $S \in J$. Therefore, the regression sum of squares can be written

$$\hat{y}'\hat{y} = \sum_{S \in J} \hat{y}'M[S]\hat{y} + \hat{y}'(I - \sum_{S \in J} M[S])\hat{y}$$

$$= \sum_{S \in J} y'M[S]y + \text{remainder}.$$
For the immediate discussion, the second term on the right side of expression 7.44 will be called a "remainder" and will be examined in more detail later. The sums of squares \( y' M[S] y \) will be displayed in a different fashion from the standard analysis of variance table. This will help to illustrate the partial orthogonalities introduced earlier.

For example, consider a three-factor main-effects model

\[
(7.45) \quad y = X_1 \beta_1 + X_2 \beta_2 + X_3 \beta_3 + \epsilon.
\]

The analysis of variance can be displayed as in Table 7.4.
The lines drawn in Table 7.4 indicate the partial orthogonalties always present in the analysis. Two sums of squares in this table are independently distributed if they can be joined by following the lines either up the table or down the table but not in both directions. The degrees of freedom associated with any of the sums of squares \( y'M[S]y \) in Table 7.4 is rank \((M[S])\).

Table 7.5 displays this type of analysis for a four-factor main-effects model. Again, entries that can be joined by following the lines up in the table or down in the table but not in both directions are independent. For the large majority of cases many of the spaces represented in Table 7.5 will be null; and therefore, the corresponding sum of squares will not appear.

It is interesting to note the type of analysis obtained for a factorial model with interactions. Table 7.6 illustrates the analysis of a three-factor experiment with all interactions present in the model. The model can be written

\[
y = \sum_{i=1}^{7} x_i \beta_i,
\]

where \( \beta_1, \beta_2, \beta_3 \) correspond to the main-effect parameters for the three factors; \( \beta_4, \beta_5, \beta_6 \) correspond to the three two-factor interactions; and \( \beta_7 \) corresponds to the three-factor interaction. The inherent nesting of the interactions forms a superstructure above the main-effects that is the mirror image of the structure below the main-effects. Notice that the analysis below the dashed line is in fact the main-effect analysis.
Table 7.5. A four-factor main-effects model.

\begin{align*}
F(1); y'M[1]y &\quad \text{Confounding} \\
F(2); y'M[2]y &\quad \text{Confounding} \\
F(3); y'M[3]y &\quad \text{Confounding} \\
F(4); y'M[4]y &\quad \text{Confounding}
\end{align*}

\begin{align*}
F(1) + F(2) &\quad y'M[1,2]y \\
F(1) + F(3) &\quad y'M[1,3]y \\
F(1) + F(4) &\quad y'M[1,4]y \\
F(2) + F(3) &\quad y'M[2,3]y \\
F(2) + F(4) &\quad y'M[2,4]y \\
F(3) + F(4) &\quad y'M[3,4]y
\end{align*}

\begin{align*}
F(1) + F(2) + F(3) &\quad y'M[1,2,3]y \\
F(1) + F(2) + F(4) &\quad y'M[1,2,4]y \\
F(1) + F(3) + F(4) &\quad y'M[1,3,4]y \\
F(2) + F(3) + F(4) &\quad y'M[2,3,4]y
\end{align*}

\begin{align*}
F(1) + F(2) + F(3) + F(4) &\quad y'M[1,2,3,4]y
\end{align*}

\begin{align*}
\text{ERROR} \\
y'R[1,2,3,4]y
\end{align*}
Table 7.6. A three-factor experiment with interactions.

Three-factor Interaction
F(1), F(2), and F(3)
y'M[7]y

Two-factor Interaction
F(1) and F(2)
y'M[1,4,7]y

Two-factor Interaction
F(1) and F(3)
y'M[5,7]y

Two-factor Interaction
F(2) and F(3)
y'M[6,7]y

Factor F(1)
y'M[1,4,5,7]y

Factor F(2)
y'M[2,4,6,7]y

Factor F(3)
y'M[3,5,6,7]y

Confounding
F(1) + F(2)
y'M[3]y

Confounding
F(1) + F(3)
y'M[2]y

Confounding
F(2) + F(3)
y'M[1]y

Confounding
F(1) + F(2) + F(3)
y'M[1,2,3,4,5,6,7]y

ERROR
y'R[7]y
F. Relationship to More Standard Analyses

The analyses of variance in Tables 7.4, 7.5 and 7.6 do not in general give the same sums of squares produced by more common procedures. For a main-effects model the only sums of squares that can be obtained by more conventional methods are those corresponding to $y'M[i]y$, $i = 1, \ldots, p$. These can be interpreted as the reduction due to fitting $X_i$ after fitting $X_1 \ldots X_{i-1} X_{i+1} \ldots X_p$. In this section a method is illustrated whereby the analysis of the form in Table 7.7 can be obtained.

Table 7.7. Sequential reduction analysis.

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Sums of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to fitting $\beta_1$</td>
<td>$\text{rank}(X_1)$</td>
<td>$y'P[1]y$</td>
</tr>
<tr>
<td>Due to fitting $\beta_2$ after $\beta_1$</td>
<td>$\text{rank}(X_1X_2) - \text{rank}(X_1)$</td>
<td>$y'(P[1,2] - P[1])y$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Due to fitting $\beta_p$ after $\beta_1 \ldots \beta_{p-1}$</td>
<td>$\text{rank}(X_1 \ldots X_p) - \text{rank}(X_1 \ldots X_{p-1})$</td>
<td>$y'(P[1, \ldots, P] - P[P])y$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - \text{rank}(X_1 \ldots X_p)$</td>
<td>$y'R[1 \ldots p]y$</td>
</tr>
<tr>
<td>Total</td>
<td>$n$</td>
<td>$y'y$</td>
</tr>
</tbody>
</table>

For the case where $p = 2$, the following analysis of variance can be constructed.
Table 7.8. The general two-factor analysis.

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor F(1)</td>
<td>Remainder</td>
<td>Factor F(2)</td>
</tr>
</tbody>
</table>

(4) Confounding
$F(1) + F(2)$
$y'M[1,2]y$

(5) ERROR
$y'R[1,2]y$

The "Remainder," term (2) in Table 7.8, can be defined to be


From Chapter IV the sum of quantities (1), (4) and (2) is the sum of squares for fitting $\beta_1$. The fact that the quantity (3) is the sum of squares for fitting $\beta_2$ after $\beta_1$ implies that an analysis identical to that in Table 7.7 can easily be constructed. Using the same "Remainder" term (2), the sum of squares due to fitting $\beta_2$ is the sum of (2), (3) and (4). Therefore, the reverse analysis can also be constructed. It is
also useful to note that the remainder $(2)$ is zero when $C(R[2]X_1)$ is orthogonal to $C(R[1]X_2)$. The converse is true with probability one; however, term $(2)$ is a function of the observations and can therefore be very small even when the spaces $C(R[2]X_1)$ and $C(R[1]X_2)$ are not orthogonal.

For the case where $p = 3$, four remainder terms can be defined. The form for such an analysis is displayed in Table 7.9.

Table 7.9. The general three-factor analysis.

\[
\begin{align*}
\text{Factor } F(1) & \quad \text{Factor } F(2) \quad \text{Factor } F(3) \\
& \quad & \\
(2) & \quad (4) & \quad (8) \\
\text{Confounding } F(1) + F(2) \quad \text{Confounding } F(1) + F(3) \quad \text{Confounding } F(2) + F(3) \\
y'M[1,2]y & \quad y'M[1,3]y & \quad y'M[2,3]y \\
& \quad & \\
(1) & \quad & \\
\text{Confounding } F(1) + F(2) + F(3) \\
y'M[1,2,3]y \\
& \quad & \\
\text{ERROR} \\
y'R[1,2,3]y
\end{align*}
\]
The four remainder terms are identified as (3), (5), (7) and (10). The first of these, (3), can be defined such that

\[(7.48) \quad y'M[1,2]y = (1) + (2) + (3).\]

In other words, the sum of (1), (2) and (3) is the result of fitting \(y\) to \(C(X_1) \cap C(X_2)\). The remainder term (5) can be defined such that the sum of (1), (2), (3), (4) and (5) gives the sum of squares of regression from fitting \(y\) to the space spanned by vectors in \(C(X_1) \cap C(X_2)\) and \(C(X_1) \cap C(X_3)\). The remainder term (7) can be defined such that

\[(7.49) \quad y'P[1]y = (1) + (2) + \ldots + (7).\]

Similarly (10) can be defined such that

\[(7.50) \quad y'P[1,2]y = (1) + (2) + \ldots + (10).\]

Using the quantities (1) through (11) defined by this procedure, the analyses represented in Table 7.7 can be constructed from Table 7.9.

There are, of course, an infinite number of ways that various remainder terms could be defined throughout a table such as Table 7.9. The above procedure could be extended to \(p = 4\), or another procedure which produces an analysis different from that of Table 7.7 could be introduced.
G. The Variances of BLUE's of Estimable Functions

Let \( J \) denote the class of sets \( S \) such that the space \( C(S) \) is non-empty. Then the variance of the function \( a'y \) with \( a \in C(S) \) is

\[
\text{(7.51)} \quad \text{Var}(a'y) = \sigma^2 a'a .
\]

Since \( a = R[S]M[S]p \) for some vector \( p \),

\[
\text{(7.52)} \quad \text{Var}(a'y) = \sigma^2 p'M[S]R[S]M[S]p .
\]

Letting the matrix \( B \) denote any basis for intersection of \( C(X_i) \) for all \( i \in S \), it follows that

\[
\text{(7.53)} \quad \text{Var}(a'y) = \sigma^2 \eta'B'R[S]B\eta
\]

for some vector \( \eta \).

In general, any estimable function \( a'y \) can be written as

\[
\text{(7.54)} \quad a'y = \sum_{S \in J} R[S]M[S]\eta[S] ,
\]

where \( \eta[S] \) is a vector dependent upon the set \( S \). The variance of \( a'y \) can be written as
(7.55) \[ \text{var}(a'y) = \sigma^2 \{ \sum_{S \in J} \eta[S]M[S]R[S]M[S]n[S] \]

\[ S \subset T \]
\[ T \not\subset S \]

The second term in expression 7.55 is essentially the sum of the covariances since the spaces \( C(S) \) and \( C(T) \) are orthogonal for \( S \subset T \) or \( T \not\subset S \).
In this chapter the decomposition theorem (Theorem 4.2) is generalized to include the case of an arbitrary covariance structure defined on the vector of random variables ε. That is, consider the linear model

\[ y = \sum_{i=1}^{p} X_i \beta_i + \varepsilon, \]

where ε is a vector of random variables with mean zero and variance \( \sigma^2 V \). The only conditions put on the matrix V are that it be real, symmetric and non-negative definite.

Zyskind and Martin (1969, page 1192) establish the following generalized Gauss-Markov theorem.

**Theorem (Generalized Gauss-Markov).** Given the linear model \( y = X\beta + \varepsilon \) with \( \text{Cov}(\varepsilon) = \sigma^2 V \), where V is any known nonnegative matrix, a non-empty subclass \( V \), dependent on the relation between X and V, of the class of all conditional inverses of V can be constructed so that for any estimable \( \lambda'\beta \) and any \( V^* \) in \( V \), a b.l.u.e. of \( \lambda'\beta \) is given by \( \lambda'\hat{\beta} \), where \( \hat{\beta} \) is any solution to the general normal equations (G.N.E.)

\[ X'V^*X = X'V^*y. \]

We will apply this theorem to the partitioned model 8.1, first with \( p \) equal to two and then for any \( p \).

Zyskind (1967) has shown that a linear function of the observations \( w'y \) is BLUE for its expectation if and only if the vector \( Vw \) is contained in \( C[X] \). From Corollary 1.2 of Zyskind and Martin (1969), it is clear that a characterization of BLUE's of estimable functions \( \lambda'\beta \) can be made by
considering the space $C(V^*X)$. From the corollary a BLUE of any estimable function $\lambda^\prime \beta$ exists of the form $\rho^\prime X^\prime V^*y$ where $\rho$ is a solution to the generalized conjugate normal equations. From the construction of $V^*$, any vector $a$ such that $a \in C(V^*X)$ also satisfies $Va \in C(X)$; and therefore, $a'y$ is BLUE for its expectation.

The normal equations for a two-partition linear model can be written as

$$
\begin{pmatrix}
X_1^\prime V^*X_1 & X_1^\prime V^*X_2 \\
X_2^\prime V^*X_1 & X_2^\prime V^*X_2
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2
\end{pmatrix}
= 
\begin{pmatrix}
X_1^\prime V^*y \\
X_2^\prime V^*y
\end{pmatrix}.
$$

Consider the following lemma from Zyskind and Martin (1969).

**Lemma** \(\text{rank } (AB) = \text{rank } (A) - k\) if and only if

$$
\dim [C(A^\prime) \cap C(B)] = k.
$$

Using the notation in 8.3 and the above lemma, the following theorem can be proved.

**Theorem 8.1** Given the matrix $V^*$ such that

$$
\text{rank } \begin{pmatrix}
X_1^\prime V^*X_1 & X_1^\prime V^*X_2 \\
X_2^\prime V^*X_1 & X_2^\prime V^*X_2
\end{pmatrix}
= \text{rank } \begin{pmatrix}
X_1^\prime \\
X_2^\prime
\end{pmatrix}.$$

then

i) \( \text{rank}(X_1^{\top}X_1) = \text{rank}(X_1) \)

and

ii) \( \text{rank}(X_2^{\top}X_2) = \text{rank}(X_2) \).

Proof: Clearly, \( \text{rank}(X^{\top}X^*) = \text{rank}(X) \). Then from the lemma,

\[ C(X) \cap C_1(X^*) = \phi. \]

Since \( C(X_1) \subseteq C(X) \), it also follows that

\[ C(X_1) \cap C_1(X^*) = \phi, \]

and \( \text{rank}(X_1^{\top}X^*) = \text{rank}(X_1) \). By a similar argument

\[ \text{rank}(X_2^{\top}X^*) = \text{rank}(X_2). \]

Therefore, \( C(X_1^{\top}X^*) = C(X_1) \). This implies that there exists a nonsingular matrix \( B \) such that

\[ X_1^{\top}X^* = X'B. \]

Postmultiplying the above expression by \( X_1 \) gives

\[ X_1^{\top}X_1 = X_1'BX_1. \]

Since \( B \) is nonsingular, it follows that

\[ \text{rank}(X_1^{\top}X_1) = \text{rank}(X_1). \]

Similarly,

\[ \text{rank}(X_2^{\top}X_2) = \text{rank}(X_2). \]
Theorem 8.1 establishes the existence of a matrix $M_1$ such that $M_1X'_1V'*X'_1 = X'_1$. Multiplying the first row partition of expression 8.3 by $-X'_2V*M'_1$ and adding the result to the second row partition gives the generalized reduced normal equations,

$$
(X'_2V*X_2 - X'_2V*M'_1X'_1V*X_2)\beta_2 = (X'_2V* - X'_2V*M'_1X'_1V*)y.
$$

Letting $M'_1X'_1V* = Q_1$ and $R_1 = I - Q_1$ gives

$$
X'_1V*R_1X_1\beta_2 = X'_2V*R_1y.
$$

Clearly, $Q_1X_1 = X_1$ and $R_1X_1 = \phi$. Consider any vector $a \in C(R'_1V'*X_2)$. There exists a vector $\rho$ such that

$$
a = R'_1V'*X_2\rho,
$$

and

$$
Va = V[I - V'*X'_1M'_1]V'*X_2\rho.
$$

Combining terms,

$$
Va = VV*[I - X'_1M'_1V'*]X_2\rho,
$$

and because of the nature of $VV'$, the vector $Va$ is contained in $C(X)$. From this it is known that $a'y$ is BLUE for its expectation, which is
(8.6) \[ E[a'y] = \rho'X_1'VR_1(x_1\beta_1 + x_2\beta_2) \]
\[ = \rho'X_1'VR_1x_2\beta_2 \]
\[ = \lambda_2^2\beta_2. \]

This establishes what may be called the generalized reduced conjugate normal equations for \( \beta_2 \) eliminating \( \beta_1 \),

(8.7) \[ X_1'R_1'X_2\rho = \lambda_2 . \]

The above development can be repeated obtaining the generalized reduced normal equations for \( \beta_1 \) eliminating \( \beta_2 \),

(8.8) \[ X_1'V'R_2X_2\beta_1 = X_1'V'R_2\gamma , \]

where \( M_2X_1'V'X_2 = X_2, Q_2 = M_2X_1'V' \), and \( R_2 = I - Q_2 \). Any vector \( a \) such that

(8.9) \[ a \in C[R_2'V'*X_2] \]

satisfies

(8.10) \[ a = R_2'V'*X_2\rho \]

for some vector \( \rho \), and

(8.11) \[ V\alpha = VV'*[I - X_2'V'*]X_2\rho . \]
This implies \( a'y \) is BLUE for its expectation. Since \( R_2X_2 = \emptyset \), the expected value of \( a'y \) is

\[
E[a'y] = \rho_1'X_1'\nu*R_2'X_2\beta_1
= \lambda_1'\beta_1 .
\]

Consider any vector \( a \) such that

\[
a \in [C(\nu'X_1) \cap C(\nu'X_2)] .
\]

Clearly, \( a'y \) is BLUE for its expectation, and there exist vectors \( \eta_1 \) and \( \eta_2 \) such that

\[
a = \nu'X_1\eta_1
= \nu'X_2\eta_2 .
\]

The expected value of \( a'y \) is

\[
E[a'y] = \eta_1'X_1'\nu*X_1\beta_1 + \eta_1'X_1'\nu*X_2\beta_2
= \eta_2'X_2'\nu*X_1\beta_1 + \eta_2'X_2'\nu*X_2\beta_2
= \lambda_1'\beta_1 + \lambda_2'\beta_2 .
\]

Any vector \( a \) contained in \( C(\nu'X_2) \) has the general form

\[
a = R_1'\nu*X_2\rho ,
\]
for some vector \( \rho \). Since \( R'_1 = I - V'^*X'_1M'_1 \), \( a \) must be a linear combination of the columns of \( V'^*X'_1 \) and \( V'^*X'_2 \). Therefore \( a \in C(V'^*X) \).

Similarly, any vector contained in \( C(R'_2 V'^*X'_1) \) must also be a vector in \( C(V'^*X) \). In addition, any vector contained in \( [C(V'^*X'_1) \cap C(V'^*X'_2)] \) is also contained in \( C(V'^*X) \). It can therefore be stated that any vector \( a \) made up of a linear combination of vectors from the three spaces \( C(R'_2 V'^*X'_1), C(R'_1 V'^*X'_2), \) and \( [C(V'^*X'_1) \cap C(V'^*X'_2)] \) must also satisfy

\[
a \in C(V'^*X) .
\]

Now consider any vector \( a \) such that \( a \in C(R'_1 V'^*X'_2) \) and \( a \in [C(V'^*X'_1) \cap C(V'^*X'_2)]. \) There exist vectors \( \eta_1 \) and \( \eta_2 \) such that

\[
(8.15) \quad a = V'^*X'_1 \eta_1 ,
\]
\[
= R'_1 V'^*X'_2 \eta_2 .
\]

This gives

\[
(8.16) \quad a = R'_1 R'_1 V'^*X'_2 \eta_2
\]
\[
= R'_1 V'^*X'_1 \eta_1
\]
\[
= \phi .
\]

Therefore, the spaces \( C(R'_1 V'^*X'_2) \) and \( [C(V'^*X'_1) \cap C(V'^*X'_2)] \) are disjoint; similarly \( C(R'_2 V'^*X'_1) \) and \( [C(V'^*X'_1) \cap C(V'^*X'_2)] \) are disjoint.

Assume there exists a vector \( a \) such that
(8.17) \[ a \in C(R^V \cdot X^\circ) , \quad a \in C(R_1^V \cdot X_2) . \]

This gives

(8.18) \[ a = R^V_2 X_1 \eta_1 = R^V_1 X_2 \eta_2 \]

for some vectors \( \eta_1 \) and \( \eta_2 \). From 8.18 it follows that \( Q_1 a = Q_2 a = \phi \).

If our assumption is correct, then

(8.19) \[ \text{rank}(X^V \cdot R_2) = \text{rank} \begin{pmatrix} X^V \cdot R_2 \\ X^V \cdot R_1 \end{pmatrix} \]

That is, the rank of \( X^V \cdot R_2 \) is the same as when it is augmented by \( a' \).

We know that

(8.20) \[ \text{rank}(X^V \cdot R_2) = \text{rank}(X^V) - k_1 , \]

where \( k_1 = \text{dim}[C(V^* X_1) \cap C(Q_2)] \). Since \( a = R^V_2 X_2 \eta_2 \), it follows that

(8.21) \[ \text{rank} \begin{pmatrix} X^V \cdot R_2 \\ X^V \cdot R_1 \end{pmatrix} = \text{rank} \begin{pmatrix} X^V \\ X^V \cdot R_1 \end{pmatrix} - k_2 , \]
where $k_2 = \dim[ C(V^*X_1, R'_1V^*X_2) \cap C(Q_2) ]$. Since $R'_1V^*X_2$ is orthogonal to $Q_2$, we have

(8.22) \hspace{1cm} k_1 = k_2 .

This gives

(8.23) \hspace{1cm} \text{rank}(X'_1V^*) = \text{rank} \begin{bmatrix} X'_1V^* \\ \eta^1_2V^*R'_1 \end{bmatrix} .

From a previous argument we know the spaces $C(V^*X_1)$ and $C(R'_1V^*X_2)$ are disjoint. Therefore, expression 8.23 holds only if

(8.24) \hspace{1cm} R'_1V^*X_2 \eta_2 = \phi .

This implies that $\phi$ is the zero vector and the spaces $C(R'_1V^*X_1)$ and $C(R'_1V^*X_2)$ are disjoint.

Thus far we have shown that the three spaces $C(R'_2V^*X_1)$, $C(R'_1V^*X_2)$ and $[C(V^*X_1) \cap C(V^*X_2)]$ are disjoint and any vector made up of a linear combination of vectors from these spaces is contained in $C(V^*X_1, V^*X_2)$. We will now demonstrate that

(8.25) \hspace{1cm} \dim[ C(V^*X_1, V^*X_2) ] = \dim[ C(R'_2V^*X_1) ] + \dim[ C(R'_1V^*X_2) ] + \dim[ C(V^*X_1) \cap C(V^*X_2) ] .
Let \( O^q \) be a basis for \( C(V^*X_1^1) \cap C(V^*X_2^1) \) and let \( O^l \) be the extension of \( O^q \) such that \( (O^l, O^q) \) forms a basis for \( C(V^*X_1^1) \). It follows that there exist matrices \( D_1 \) and \( D_2 \) such that

\[
(8.26) \quad (O^l, O^q) = V^*X_1^1 D_1
\]

and

\[
(8.27) \quad V^*X_1^1 D_2 = (O^l, O^q) .
\]

Letting \( q_1 = \dim[C(V^*X_1^1)] \) and \( q_{12} = \dim[C(V^*X_1^1) \cap C(V^*X_2^1)] \),

\[
(8.28) \quad \text{rank}(O^l) = q_1 - q_{12} ,
\]

and

\[
(8.29) \quad \text{rank}(O^q) = q_{12} .
\]

From expressions 8.26 and 8.27 it follows that \( R^l_2(O^l, O^q) \) forms a basis for \( R^l_2 V^*X_1^1 \). Since \( O^q \) is made up of vectors from \( C(V^*X_2^1) \),

\[
(8.30) \quad R^l_2(O^l, O^q) = (R^l_2 O^l_1, \phi) .
\]

The columns of \( R^l_2 O^l_1 \) can be shown to be linearly independent as follows. Assume the columns are linearly dependent. Then there exists a vector \( \rho \) such that
From the definition of $R_2'$ it follows that
\[ O_1'\rho = V^*X_2M_2'O_1'\rho . \]

Then vector $O_1'\rho \in [C(V^*X_1) \cap C(V^*X_2)]$ and $O_1'\rho \neq \phi$, but this is impossible because of the construction of $O_1$ and $O_2$. Therefore

(8.32) \[ \text{rank}(R_2'V^*X_1) = q_1 - q_{12} . \]

By a similar argument it follows that

(8.33) \[ \text{rank}(R_2'V^*X_2) = q_2 - q_{12} . \]

This gives

(8.34) \[ \text{dim}[C(R_2'V^*X_2)] + \text{dim}[C(R_2'V^*X_2)] \\
+ \text{dim}[C(V^*X_1) \cap C(V^*X_2)] \\
= q_1 + q_2 - q_{12} \\
= \text{dim}[C(V^*X_1, V^*X_2)] . \]

The above development establishes the following extension to Theorem 4.2.

**Theorem 8.2** Given the notation above, any linear function $a'y$ that is BLUE for its expectation with $a \in C(V^*X)$ has the following unique decomposition
(8.35) \[ a'y = \beta_1'y + \beta_2'y + \beta_3'y, \]

where \( \beta_1 \) is a vector in \( \mathbb{C}(\mathbb{R}^{V'X_1}) \), \( \beta_2 \) is a vector in \( \mathbb{C}(\mathbb{R}^{V'X_2}) \), and \( \beta_3 \) is a vector in \( \mathbb{C}(\mathbb{R}^{V'X_3}) \). In other words, \( a'y \) can be decomposed into the sum of a function that is BLUE of an estimable function of \( \beta_1 \) alone, a function that is BLUE of an estimable function of \( \beta_2 \) alone, and a function that is BLUE of an estimable function of \( \beta_1 \) and \( \beta_2 \) that cannot be further decomposed into estimable functions of \( \beta_1 \) alone and \( \beta_2 \) alone.

Using Theorem 8.2 the sums of squares for regression \( \hat{\beta}'X'V'X\hat{\beta} \) can be decomposed in a manner similar to that of Chapter IV. That decomposition as well as variances of BLUE's, interfactor information, and their multifactor extensions will not be developed here.
IX. A COMPUTATIONAL PROCEDURE

In this chapter some computational aspects of the theory presented in previous chapters are discussed. Much has been written [see Muller and Wilkinson (1971)] concerning the computational aspects of the analysis of variance. However, none of the presently available procedures provide the type of decomposition introduced in Chapter IV and Chapter VII. Such a procedure will be discussed here.

A. Analysis of Variance Programs

Most computer programs can be thought of as consisting of three basic parts, the input to the program, the algorithm to perform the computations, and the output of the results. The amount of communication between these parts varies considerably as in Figure 9.1 for a batch processing mode and Figure 9.2 for a time-sharing mode.

![Figure 9.1. Batch processing mode](image)

Nearly all analysis of variance programs are implemented in a batch processing mode. This is natural, since most computers operate in a batch processing mode. However, time-sharing versions of analysis of
Figure 9.2. Time-sharing mode

'variance programs also exist. The author has implemented such a program using the extension of Yates' algorithm on the time-sharing system at Iowa State University.

The input to an analysis of variance program generally consists of some form of model specification, physical characteristics of the experiment such as the number of levels of each factor, and perhaps some type of information as to what hypotheses are to be tested.

The algorithm generally operates on the data using a series of regressions, linear combinations of means, or other operations intended to evaluate the needed expressions. In any program the algorithm is very important as it determines the usefulness of the program as a whole. This chapter is concerned with the algorithm for an analysis of variance program.
The output from any analysis of variance program generally consists of an analysis of variance table, with the option available at the input stage to produce such information as means, plots of residuals, estimates of predefined contrasts, etc. Some output may go to an external storage device to be used as input to other programs.

The report by Muller and Wilkinson (1971) contains descriptions of numerous methods for specifying the model, algorithms, and forms for output.

B. Constructing A Basis for the Intersection of Two Column Spaces

The intersection of two column spaces clearly forms the foundation for a general analysis of variance program. In this section a method for constructing a basis for the intersection of two column spaces will be described. This method employs a procedure for calculating a generalized inverse. There are two reasons for using such a procedure. First, methods for calculating generalized inverses usually provide the information concerning linear relationships among the columns necessary for determining a basis for the intersection space. Secondly, the resulting generalized inverse will be useful for determining parameter estimates, sums of squares, and estimability. Many procedures for calculating generalized inverses exist. A method due to Greville (1960) will be described here. This method was chosen merely because it admits a concise recursive procedure that produces the information needed.

Let $A$ be an $n \times m$ matrix and let $A_k$ represent the first $k$ columns of the matrix $A$. The generalized inverse of the first column of $A$ can be written as
where \( c_1 = 1/(A_1^t A_1) \). Then \( A_k^+ \) for \( k = 2, \ldots, m \) can be calculated by the following procedure. Let \( a_k \) denote the \( k \)-th column of \( A \). Then

\[
A_k = (A_{k-1}, a_k).
\]

Define

\[
d_k = A_{k-1}^t a_k,
\]

\[
c_k = a_k - A_{k-1} d_k,
\]

and

\[
b_k = c_k^t \quad \text{if} \quad c_k \neq \phi,
\]

\[
b_k = d_k^t A_{k-1}^t / (1 + d_k^t d_k), \quad \text{if} \quad c_k = \phi.
\]

Then

\[
A_k^+ = \begin{bmatrix} A_{k-1}^+ - d_k b_k \\ b_k \end{bmatrix}.
\]

Notice that if \( c_k = \phi \) then \( a_k \) is a linear combination of the first \( k-1 \) columns of \( A \) and the coefficients that define the linear
combination are the elements of vector $d_k$. A flowchart for the Greville method is shown in Figure 9.3.

Such a procedure can now be used to construct a basis for the intersection of two column spaces. Let $Z_1$ and $Z_2$ be $m \times p_1$ and $m \times p_2$ matrices, respectively. Let $z_{ij}$ denote the $j$-th column of $Z_i$. A procedure to construct a basis for $\mathcal{C}(Z_1) \cap \mathcal{C}(Z_2)$ will be described as follows. Calculate the generalized inverse of $Z_1$, denoted by $Z_1^\dagger$. Use the Greville method to calculate

$$
(9.7) \quad (z_1, z_{21})^\dagger.
$$

If $z_{21}$ is a linear combination of the columns of $Z_1$ then $z_{21}$ is a vector in $\mathcal{C}(Z_1) \cap \mathcal{C}(Z_2)$. Let $b_1 = z_{21}$ be the first basis vector. If $z_{21}$ is not a linear combination of the columns of $Z_1$, then no information concerning the basis is available at this stage. Next calculate

$$
(9.8) \quad (z_1, z_{21}, z_{22})^\dagger.
$$

If $z_{22}$ is a linear combination of the columns of $(Z_1, z_{21})$, say $z_{22} = a_1 z_{21} + Z_1 d$ for some scalar $a_1$ and some vector $d$, then the vector

$$
(9.9) \quad b = z_{22} - a_1 z_{21}
$$

is a vector in $\mathcal{C}(Z_1) \cap \mathcal{C}(Z_2)$. If this is not the first vector in the basis, then compute
Figure 9.3. Flowchart of the Greville method
In this manner an independent set of basis vectors can be constructed. If $b$ is the first vector in the basis then let $b_1 = b$. Regardless of whether or not $z_{22}$ is a linear combination of the columns of $(z_1, z_{21})$, the next step is to calculate

\[(9.10) \quad (z_1, z_{21}, z_{22}, z_{23})^\dagger.\]

For the $k$-th step, $k = 1, 2, \ldots, p_2$, let $B_k$ denote the matrix of independent basis vectors, let $Z_{2k}$ denote the first $k$ columns of the matrix $Z_2$ and if $z_{2k}$ is a linear combination of the columns of $Z_1(z_{2k-1})$, then denote the linear combination by

\[(9.11) \quad z_1a_1 = z_{2k}a_k,\]

for vectors $a_1$ and $a_k$. Then the procedure at the $k$-th step can be outlined as follows:

i) If vectors $a_1$ and $a_2$ exist such that 9.11 holds, then add the vector $Z_{2k}a_k$ to the basis and determine if the basis vectors are linearly independent. If the vectors are independent, add $Z_{2k}a_k$ to the basis, if not then delete $Z_{2k}a_k$. Go to step iii).

ii) If the vectors $a_1$ and $a_2$ do not exist such that 9.11 holds, then go to step iii).

iii) Increment $k$ by 1 and repeat the above process until $k = p_2$. 

\[(b_1, b)^\dagger.\]
The above procedure will yield \( \text{dim}[C(z_1)], \text{dim}[C(z_1) \cap C(z_2)], \) and \( \text{dim}[C(z_1, z_2)] \). In addition the matrices

\[
\begin{align*}
(9.12) & \quad z_1^+, \\
(9.13) & \quad (z_1, z_2)^+,
\end{align*}
\]

and

\[
(9.14) \quad B^+,
\]

where \( B \) denotes the matrix of basis vectors, are computed.

A flow diagram of the above procedure is presented in Figure 9.4.

C. The Two-factor Model

As in previous chapters, the simple case will be examined first. Therefore consider the linear model

\[
(9.15) \quad y = x_1 \beta_1 + x_2 \beta_2 + \varepsilon.
\]

The corresponding normal equations are

\[
(9.16) \quad \begin{bmatrix} x_{11}^t & x_{12}^t \\ x_{21}^t & x_{22}^t \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} x_{1y}^t \\ x_{2y}^t \end{bmatrix}.
\]
Figure 9.1. A method for constructing the basis for an intersection space
Wherever possible we will try to work with matrices no larger than \( X'X \). The matrices \( X'X \) and \( X'y \) can be constructed when the data are read. In this way matrices of the size of \( X \) need not be stored in core.

From the fact that if \( X'Xp = \phi \) for some vector \( p \) then \( Xp = \phi \), it follows that linear relationships among the columns of \( X'X \) reflect identical relationships among the columns of \( X \). Therefore, applying the procedure described in the previous section, the matrices

\[
\begin{bmatrix}
  X_1'X_1 & X_1'X_2 \\
  X_2'X_1 & X_2'X_2
\end{bmatrix}^+,
\]

(9.17)

\[
x_1^+,
\]

(9.18)

can be evaluated. In addition, matrices \( C \) and \( B \) can be constructed such that if \( B \) denotes a matrix whose columns form a basis for

\[
c \begin{bmatrix}
  X_1'X_1 \\
  X_1'X_2
\end{bmatrix} \cap c \begin{bmatrix}
  X_1'X_2 \\
  X_2'X_2
\end{bmatrix}
\]

then

\[
B = \begin{bmatrix}
  X_1'X_2 \\
  X_2'X_2
\end{bmatrix} c.
\]

(9.19)
It follows then that the columns of the matrix

\[(9.20) \quad D = X_2C\]

form a basis for \( C(X_1) \cap C(X_2) \). The sums of squares corresponding to the completely confounded effects can be evaluated by calculating

\[(9.21) \quad y'X_2C(C'X_2X_2C)^-C'X_2y . \]

The sums of squares attributable to \( \beta_1 \) alone and \( \beta_2 \) alone can be calculated by evaluating

\[(9.22) \quad \hat{\beta}_1'(x_1'x_1 - x_1'x_2(x_2'x_2)^-x_2'x_1)\hat{\beta}_1 , \]

and

\[(9.23) \quad \hat{\beta}_2'(x_2'x_2 - x_2'x_1(x_1'x_1)^-x_1'x_2)\hat{\beta}_2 , \]

respectively. The "remainder" term can be evaluated by subtraction.

Estimability of redefined linear functions of the parameters \( \lambda'\beta \) can be determined by using the fact that the product

\[(9.24) \quad x'x(x'x)^\dagger \]

is the orthogonal projection operator on \( C(x') \). Therefore \( \lambda'\beta \) is estimable if and only if

\[ \lambda - x'x(x'x)^\dagger\lambda = \phi . \]
D. Models With More Than Two Factors

The procedures introduced in the preceding section can be extended to models with more than two partitions. It will be assumed that the matrices $X'X$ and $X'y$ have been calculated and are available. The general approach will be to calculate $(X'X)^\dagger$ and at the same time construct bases for the intersections of the appropriate column spaces. These bases will not explicitly be obtained. Instead matrices $B[i_1, \ldots, i_k]$ will be obtained such that

$$XB[i_1, \ldots, i_k]$$

will form an independent set of basis vectors for

$$\bigcap_{j=1}^k C(X_{i_j}).$$

Writing $X'X$ as

$$X'X = (X'X_1, X'X_2, \ldots, X'X_p)$$

and applying the algorithm introduced in the previous section to any two partitions $X'X_i$ and $X'X_j$ the matrix $B[i, j]$ can be constructed such that

$$X_i B[i, j]$$

forms a basis for $C(X_i) \cap C(X_j)$. A basis for

$$C(X_i) \cap C(X_j) \cap C(X_k)$$
can be constructed by operating on the two matrices

\[ X'X_iB[i,j] \] and \[ X'X_k \].

This will give \( B[i,j,k] \) and

\[ X_iB[i,j,k] \]

will be a set of basis vectors for \( C(X_i) \cap C(X_j) \cap C(X_k) \). This can be
continued until all of the necessary intersections are determined. The
relevant sums of squares can be calculated by fitting each model

\[ R[\bar{s}]X_iB[S]r = y \]

for some \( i \in S \). This would at first glance appear to be a very complex
model to fit. However when the normal equations are written in the proper
form, the model is in fact simple to fit. Consider the normal equations

\[ B'[S]X_i'R[\bar{s}]X_iB[S]r = B'[S]X_i'R[S]y \]

substituting \( I - X[S](X'[S]X[S])^{-1}X[S] \) for \( R[S] \) where \( X[S] \) represents
the matrix \( X \) with all \( X_i \) such that \( i \in S \) deleted, gives


as the coefficient matrix and is generally a very small matrix. The
quantities

\[ B[S]X_iX_i \] and \( B[S]X_iX[\bar{s}] \)
are easily retrieved from calculations done during the determination of $B[\delta]$. The matrix
\[(X'[\delta]X[\delta])^\dagger\]
can also be calculated as the intersections are being determined. The right-hand sides of the normal equations can be computed similarly. Predefined contrasts in the parameters $\lambda'[\beta]$ can be checked for estimability by calculating
\[X'X(X'X)^{\dagger}\lambda \]
as in the previous section. Then if $\lambda'[\beta]$ is estimable
\[X'X(X'X)^{\dagger}\lambda = \lambda \]

Using the above procedures, all of the information discussed in Chapter VII can be computed. The most difficult problem yet to be solved will be to determine the most efficient way to implement these procedures on a computer.
X. SUMMARY

This thesis has addressed itself to some of the complex questions arising from the fitting of classificatory linear models to arbitrary data structures. In order to develop a computer algorithm for analysing arbitrary data-model situations, it is first necessary to determine the basic parts of such an analysis and then determine a useful way of presenting them. To determine the nature and characteristics of these basic parts it is necessary to examine in some detail the concepts of orthogonality, rank, confounding and estimation as they apply to the general linear hypothesis.

A. Estimability

The approach used in this thesis has been to let estimability considerations direct the analysis of a general linear classificatory model. This results in a partitioning of $C(X)$ into subspaces, each corresponding to a particular group of estimable functions. Each group of estimable functions exhibits a certain type of confounding. The concept of orthogonality in a linear model can then be identified with the orthogonality (or non-orthogonality) of the above subspaces. The partitioning of $C(X)$ also suggests a partitioning of the total sum of squares into meaningful components for testing various hypotheses. The appropriate degrees of freedom can be obtained from the dimensions of the above subspaces.

The general two-factor model with a general covariance structure $V$ is also examined. The appropriate partitions of the space $C(V'X)$ are given.
B. Comparison With More Classical Procedures

The approach to fitting general linear classificatory models seems to be that outlined in Table 7.7. The fact that this approach is inadequate for general linear classificatory models should be clear from the results of Chapter VII. In an incomplete three-factor factorial model, for example, the projection of $y$ onto some of the subspaces representing BLUE's of confounded effects cannot be calculated without first determining a basis for the intersection of appropriate column spaces. In some submaximal rank situations the effect of the factors involved can be separated into effects attributable to that factor alone and effects attributable to a combination of factors that cannot be separated. The effects attributable to each factor alone can be determined by using an approach similar to that in Table 7.7. However, the effect of confounded functions cannot be calculated by using such an approach. The most difficult problem, of course, is determining the confounded effects and corresponding intersections of column spaces.

C. Further Work

The largest area in need of further work is the area of the computer algorithm. The procedure described here can tend to be quite time consuming and may be subject to round-off errors. It is the author's belief that there is room for improvement in all of the "general" algorithms proposed in the literature since not all of the information available is used. None of the present algorithms use the fact that the model matrix is made up entirely of zeros and ones (at worse zeros, ones and negative ones).
Some computer programs make use of this fact when allocating storage, but none specifically make use of it in the algorithm itself.

Another topic for further study is in the area of combining the sums of squares in a meaningful way when the spaces $C(S)$ introduced in Chapter VII are not orthogonal. One example was given providing the construction of the analysis given in Table 7.7. It may be possible to introduce additional remainder terms so that the analysis corresponding to any ordering of the $\beta$'s can be constructed.

The subject of combining information from several experiments, though not specifically mentioned in this thesis, is closely related. In order to combine estimates of estimable functions from several experiments in the most efficient manner, it is necessary to determine the relationships between estimable functions in each of the experiments.

It would seem that perhaps the model and data structure are not sufficient to completely describe an experiment. Perhaps in addition to a statement of the model it is necessary to have a statement of the assumptions one is prepared to make in the event some confounding has occurred due to missing observations. One can imagine incorporating statements such as: "In the event that $a_1 + a_3$ is confounded with $b_2 + b_4$, there is substantive evidence for assuming that $a_1 + a_3$ is zero." The need for such statements obviously occurs with fractional factorials, but in general the interconfounding in a given data-model situation will be unknown without examination of the sort outlined in this thesis.
XI. ACKNOWLEDGEMENTS

In any work such as this there are always those whose assistance is deeply appreciated.

My parents have continually been a source of encouragement to me.

I am indebted to Dr. Dave Jowett for his guidance during several years at Iowa State University.

I would like to express my gratitude to Professor Dave Cox and Professor George Zyskind for many interesting and helpful discussions.

It has been my good fortune to work under such an able Statistician and Scientist as Professor Oscar Kempthorne. I will long remember our association.

This work was supported in part by NSF Grant GP-24614.

Without the expert assistance of Marlene Sposito who typed this manuscript this thesis would have taken even longer to complete.

Finally, my deepest gratitude to my wife, Mary Ann, who has given all of her time and energy to assist me.
XII. BIBLIOGRAPHY


Elston, R. C. and Bush, N. 1964. The hypothesis that can be tested when there are interactions in an analysis of variance model. Biometrics 20: 681-698.


